Supplemental methods

Supplementary methods

Genomic sequence data generation

Genomic DNA was isolated from peripheral blood monocytes using Gentra PureGene blood kit from Qiagen Inc. (cat# 158467). For all study subjects, genomic DNA was sequenced using 101 base-pair paired-end reversible terminator massively parallel sequencing on the HiSeq 2000 instrument at Illumina, Inc (San Diego, CA) following sequencing library preparation according to standard Illumina protocols.[1] The median insert size for paired end sequencing was 250 bp. Confirmatory sequencing of genomic DNA for nine study subjects was performed using 70 base pair DNA nanoball array and combinatorial probe-anchor ligation methods at Complete Genomics, Inc (Mountain View, CA) following sequencing library preparation according to standard Complete Genomics protocols.[2]

Sequence alignment, genetic variant and clinical genotype identification

Single nucleotide (SNV) and small insertion deletion (indel) variant identification were performed using the general framework adapted from the HugeSeq pipeline.[3] Sequence reads generated by Illumina Inc, were aligned using BWA version 0.6.1[4] to reference genomes representing the most common allele at hg19 coordinate positions for the European Ancestry in Utah (CEU) or Han Chinese from Beijing and Japanese from Tokyo (CHB/JPT) HapMap population groups, depending on self-reported ethnicity ("major allele reference genomes"). We have previously shown that alignment and variant identification using a major allele reference genome improves alignment sensitivity and accuracy of variant calls and rare variant identification.[5] PCR duplicate identification was performed using Picard tools v1.64. Local realignment around discovered and known indels (according to dbSNP 135) was performed using the Genome Analysis Toolkit (GATK) v.1.6-9[6] with a LOD threshold of 5. Base quality score

recalibration was performed using GATK v1.6-9 using the following covariates: ReadGroupCovariate, QualityScoreCovariate, CycleCovariate, DinucCovariate. The GATK unified genotyper was used to call SNVs and indel variants from the major allele reference. Variant quality score recalibration was performed using GATK v1.6-9 and the following annotations trained on HapMap 2 and 3 genotypes,[7] 1000 genomes phase 1 sequence, genotype and indel data,[8] and a curated set of indels cataloged by Mills and Devine[9]: quality by depth, homopolymer run, mapping quality rank sum, haplotype score, read position rank sum, and Fisher's exact test for strand bias.

Structural variants were identified by consensus of two or more of the following six algorithms: BreakSeqLite-1.0,[10] BreakDancer-1.1,[11] CNVnator-0.2.2,[12] Pindel-r19,[13] Delly-0.0.11,[14] and readDepth-0.9.8.4.[15] For all programs, raw calls with length shorter than 50bp were removed. For readDepth, the thresholds of copy number changes on X and Y chromosomes were adjusted to ploidy of one for male participants. For Delly, calls with mapping quality of zero were removed, and if the length of the reported SV was more than 10,000 bp, the number of supporting pairs was required to > 30 to minimize false positives. A reciprocal overlap of > 50% was required to suggest algorithm consensus.

Sequence alignment, assembly, and variant identification for Complete Genomics raw data was performed by Complete Genomics, Inc. according to published assembly and genetic variant identification algorithms.[2]

To provide a comprehensive survey of genotypes with inherited and complex disease risk and drug response implications, irrespective of reference allele, we called targeted genotypes from WGS data covering variant positions with these previously described associations. Clinical genotypes were identified using the –L command in the GATK unified genotyper for interval lists representing 72,383 (males) and 72,316 (females) variants cataloged

in the Human Gene Mutation Database (HGMD)[16,17] that reside in high impact disease genes cataloged by Clinvar,[18] 118 variants associated with diabetes mellitus type 2 and coronary artery disease risk in Caucasian and East Asian populations,[19-22] and 555 variants with clinical drug response associations as cataloged by PharmGKB (http://www.pharmgkb.org/search/clinicalAnnotationList.action, database download 1/6/2013).[23]

Coverage of genomic regions corresponding to inherited disease genes was calculated using the GATK tool CallableLoci, with two thresholds: one sequence read covering target regions and ten sequence reads covering target regions. The second criterion is a conservative threshold for discovery of heterozygous variants using current short read high throughput sequencing platforms. The percentage of exonic bases in all transcripts of 2,725 (males) and 2,716 (females) inherited disease genes covered at these thresholds by reads passing quality filters for genotype determination was calculated for each subject. Subsequently the median, mean, standard deviation, minimum, and maximum percentages of exonic bases covered at these thresholds were calculated for the twelve study subjects. In this cohort, the minimum percentage of exonic bases defined the percentage of genes always covered above specific thresholds for sequence depth and percentage of exonic bases; e.g., the percentage of genes covered by ten or more sequence reads in all individuals at >99% of exonic bases is defined by this metric. The maximum percentage of exonic bases defined, in this cohort, the percentage of genes never covered above specified thresholds for sequence depth and percentage of exonic bases; e.g., the percentage of genes never covered by ten or more sequence reads at >99% of exonic bases is defined by this metric.

Genetic variant and clinical genotype annotation

Data sources used for initial sequence data annotations are presented in the **Box**. Variants used for assessment of personal risk and carrier status for inherited disease, as well as previously reported variants in genes associated with inherited disease, were annotated using these data, which includes 25 data sources used for allele frequency estimation, 3 gene effect prediction models, 7 models for predicted variant pathogenicity, 5 data sources for sequence comparison to lower animals, and 4 data sources for clinical phenotype effect prediction. A custom perl/python pipeline, extended from annovar[24] and modified to provide annotations for custom data sources, performed sequence annotation with respect to these data. These basic annotations were carried forward into the clinical disease risk and drug response prioritization pipeline.

Bioinformatic prioritization of candidate inherited disease risk variants

We used a novel heuristic for prioritization of previously reported variants in inherited disease genes, as well as rare and novel variants in inherited disease genes with no previously reported phenotypic association (Sequence to Medical Phenotypes, STMP, ashleylab.stanford.edu/tools/tools_synthetic/stmp)**.** We first reduced all variants to a set that occurs within 2,725 genes cataloged in ClinVar,[18] manually curated to exclude drug response associations and common disease susceptibility loci, and used STMP to identify variants for manual review and consideration for reporting. Variants in STMP tiers 1-3 were retained for manual review. This set includes previously unreported (PU) and rare (<1% allele frequency in an ethnically-matched population) loss of function variants (PU-tier 1); previously unreported and rare missense variants with consensus evidence for evolutionary conservation/constraint (PU-tier 2); previously unreported and rare missense variants with consensus evidence for predicted

pathogenicity (PU-tier 3); loss of function variants with previously reported disease associations (PR-tier1); rare variants with previously reported disease associations (PR-tier 2); and missense/non-frameshift variants previously reported in association with disease with an allele frequency > 1% (PR-tier 3).

As a final filter for both previously reported variants in inherited disease genes and previously unreported, rare variants in inherited disease genes, we excluded variants from manual review and consideration for reporting if they were seen more than three times (tier 1 variants) or more than two times (tiers 2 and 3) in the cohort of 12 individuals. This filter serves to identify and exclude variants whose previously unappreciated high allele frequency is a result of false negatives in population genetic surveys or systematic false positives specific to our sequence assessment pipeline. Similar filters have proven effective in excluding such systematic artifacts in other contexts.[25]

Manual review of candidate inherited disease risk variants

For each variant identified by STMP for manual review, a multi-disciplinary genomics team (MDGT) comprised of three genetic counselors, three physicians/informaticists, and one molecular pathologist generated one of nine summary statements (**Table S1**) describing the likely pathogenicity with respect to inherited disease risk and also characteristics of the phenotype for reporting purposes (Table S2). This is an extension of our previously applied framework for evaluating variants with potential inherited disease risk implications.[5] Primary literature reports of linkage, association, and functional assays were reviewed when available, and summary statements were gathered from OMIM, GeneReviews, and locus specific databases. Our internal classification scheme differs from the reporting classification scheme proposed by the American College of Medical Genetics and Genomics[26] (ACMG recommendations for standards for interpretation and reporting of sequence variations:

Revisions 2007) in order to capture additional granularity in the potential pathogenicity of previously reported and unreported variants in the context of low prior inherited disease risk probability. Following summary statement generation, we classified reportable variants into one of three categories that map to ACMG reporting classifications (full description in the **Box** in the main manuscript**;** map to variant classification in **Table S1**) or classified them as non-reportable. We considered three types of variants to be reportable: 1) variants with previously reported inherited disease associations supported by co-segregation, functional evaluation, or differential allele frequency in disease cases vs control populations (maps to current ACMG category 1: " Sequence variation is previously reported and is a recognized cause of the disorder"); 2) expected pathogenic variants that cause loss-of-function in genes in which variants of that type have previously been implicated in a inherited disease phenotype (maps to current ACMG category 2: "Sequence variation is previously unreported and is of the type which is expected to cause the disorder"); 3) reportable variants of uncertain significance (VUS), corresponding to variants commonly referred to as "VUS-favor pathogenic" (maps to current ACMG category 3: "Sequence variation is previously unreported and is of the type which may or may not be causative of the disorder"). The reportable VUSs were comprised of variants with previously reported disease associations based on conflicting evidence, and previously unreported missense and nonframeshift variants with consensus evidence of constraint/conservation or predicted pathogenicity, occurring in genes in which that variant type has been associated with disease. This last category captures variants that, in the correct family and personal medical history context, might potentially confer personal risk or carrier status for inherited disease, whereas the first two categories represent variants for which the preponderance of genetic evidence may suggest such risk independent of the clinical prior probability of disease. All variants underwent initial review by a genetic counselor or physician, and were over-read by at

least one additional observer. When disease-specific knowledge was required for assessment of pathogenicity, or when disagreement between primary reviewers and over-readers occurred, additional expert reviewer consultation was pursued. In addition to this over-reading, we assessed blind inter-rater agreement among six genomics professionals as described in the main text for 18 randomly selected variants chosen from those discovered in the study participants.

Assessment of genetic risk for coronary artery disease and diabetes

We generated percentile genetic risk scores for Caucasian subjects using genotypes for 47 and 45 common SNVs with replicated associations with diabetes mellitus type 2 and coronary artery disease, respectively (variant lists in **Tables S3-S6**).[19,21] Similar to published scores for assessment of cardiometabolic genetic risk,[27] an additive risk score was generated by summing the number of risk alleles at each locus, weighted by the published odds ratio for each risk SNV. This risk score was compared with individual risk scores generated for Caucasian subjects from the Atherosclerosis Risk in Communities 2 cohort[28] to generate the population percentile of genetic risk. A similar risk score was generated for subjects with East Asian ancestry by comparison of genetic risk scores at 18 and 8 common SNVs with replicated DM2 and CAD associations, respectively,[20,22] in East Asian cohorts to genetic risk scores from 90 CHB/JPT subjects from HapMap 2.[7]

Supplementary Tables

Box. Data sources used for whole genome sequence variant annotation.

Allele frequency estimation
National Heart Lung Blood Institute Exome Sequencing Project 6500 (all) National Heart Lung Blood Institute Exome Sequencing Project 6500 (European Americans) National Heart Lung Blood Institute Exome Sequencing Project 6500 (African Americans) HapMap 2 and 3 African ancestry in Southwest USA HapMap 2 and 3 Utah residents with Northern and Western European ancestry from the CEPH collection HapMap 2 and 3 Han Chinese in Beijing, China HapMap 2 and 3 Chinese in Metropolitan Denver, Colorado HapMap 2 and 3 Gujarati Indians in Houston, Texas HapMap 2 and 3 Japanese in Tokyo, Japan HapMap 2 and 3 Luhya in Webuye, Kenya HapMap 2 and 3 Mexican ancestry in Los Angeles, California HapMap 2 and 3 Maasai in Kinyawa, Kenya HapMap 2 and 3 Toscani in Italia HapMap 2 and 3 Yoruba in Ibadan, Nigeria 1000 genomes project pilot 1, 2010 November release, all subjects 1000 genomes project phase 1, 2011 May release, all subjects 1000 genomes project phase 1, 2012 February release, all subjects 1000 genomes project phase 1, 2012 April release, all subjects 1000 genomes project phase 1, 2012 April release, European ancestry 1000 genomes project phase 1, 2012 April release, East Asian ancestry 1000 genomes project phase 1, 2012 April release, West African ancestry 1000 genomes project phase 1, 2012 April release, American ancestry Complete Genomics public panel (cg46) Complete Genomics diversity panel (cg69) National Center for Biotechnology Information database of Single Nucleotide Polymorphisms database, version 135
Gene effect prediction
UCSC known gene NCBI RefSeq GenCode
Functional effect prediction
Likelihood Ratio Test Sorting Intolerant From Tolerant PolyPhen2 Mutation Taster Database of nonsynonymous function prediction Segmental duplication database Database of Genomic Variants
Phenotype association
Human Gene Mutation Database National Center Biotechnology Information ClinVar National Human Genome Research Institute Genome Wide Association Study catalog Pharmacogenomics Knowledge Base

Table S1. Summary statement definitions for inherited risk candidates.

Summary statement	Reporting classification		Definition
	If variant is previously described	If variant is not previously described	
Likely benign, strong evidence	Do not report	N/A	• High frequency variants – very likely too common to cause inherited disease or to have implications for carrier status, with similar frequency seen in cases and controls ---------- OR ---------- • Functional evidence consistently demonstrating variant is benign
Some evidence suggesting benign	Do not report; maintain for future curation	N/A	• The role of the variant is unclear, but the preponderance of available evidence suggests that it may be benign ---------- OR ---------- • Common allele frequency in the general population, likely too common to cause disease ---------- OR ---------- • Rare variant with no previously described disease association -- AND -- Prediction programs consistently predict tolerated or benign
Novel, predicted benign	N/A	Do not report, maintain for future curation	• Novel variant, not found in any queried datasets (NHLBI ESP, 1000Genomes, HapMap Project, Complete Genomics and others, if available) or available literature; no allele frequency data available; no entry in dbSNP -- AND -- Variant is consistently predicted benign by all *in silico* programs queried
Uncertain	Report as "variant of	If variants of this type	• Role of variant in disease as described in the literature is unclear; balanced and

significance	uncertain significance"	have been implicated in disease: Report as "variant of uncertain significance" If variants of this type have not been implicated in disease: Do not report, maintain for future curation	conflicting information in the published literature about pathogenicity -- AND -- Variant is absent or detected in control datasets at frequency that falls reasonably within the range of expected frequency and up to 2 times the expected frequency based on estimated disease prevalence ---------- OR ---------- • Variant with no previously described disease association, is absent or in control datasets at a frequency that falls reasonably within the range of expected frequency and up to 2 times the expected frequency based on estimated disease prevalence -- AND – Some evidence of conservation and predicted pathogenicity
Novel, predicted damaging by 1 program	N/A	Do not report, maintain for future curation	• Novel variant, in which one *in silico* program predicts damaging/disease-causing of available *in silico* prediction programs (Mutation Taster, Mutation Taster chromosomal position query, SIFT, SIFT indel, LRT, PolyPhen)
Novel, predicted damaging by 2 or more programs	N/A	If variants of this type have been implicated in disease: Report as "variant of uncertain significance" If variants of this type have not been implicated in disease: Do not report,	• Novel variant, in which 2 *in silico* prediction programs (or more) predict damaging/disease-causing

		maintain for future curation	
Some evidence suggesting pathogenic	Report as "variant of uncertain significance"	Report as "variant of uncertain significance"	• Role of the variant in disease is somewhat unclear but the preponderance of evidence suggests it may be disease causing -- AND – Variant is absent or detected in control datasets at frequency that falls reasonably within the range of expected frequency and up to 2 times the expected frequency based on estimated disease prevalence ---------- OR ---------- • Variant is reported in a single proband in a manner consistent with mode of inheritance, but without additional evidence -- AND – Variant is absent or detected in control datasets at frequency that falls reasonably within the range of expected frequency and up to 2 times the expected frequency based on estimated disease prevalence ---------- OR ---------- • Variant with no previously described disease association that disrupts a protein domain or residue considered to be of critical importance to normal gene function and/or often implicated in disease -- AND – Variant is absent or detected in control datasets at frequency that falls reasonably within the range of expected frequency and up to 2 times the expected frequency based on estimated disease prevalence ---------- OR ---------- • Very rare or novel loss-of-function

			variant of a type in a gene where that type of variant has been inconsistently implicated in disease
Likely pathogenic	Report as "reported disease associated mutation"	Report as "rare, expected pathogenic variant"	• Very rare or novel loss-of-function variant of a type in a gene where that type of variant has been clearly implicated in disease ---------- OR ---------- • Variant is reported in a single proband in a manner consistent with mode of inheritance -- AND -- Available functional and/ or animal model evidence suggests evidence of functionality/disruptive effect on normal gene function ---------- OR ---------- • Variant convincingly segregates with disease in one family and is seen very infrequently* in controls ---------- OR ---------- • Variant is reported in multiple unrelated probands in a manner consistent with mode of inheritance -- AND -- Variant is absent or detected in control datasets at frequency that falls reasonably within the range of expected frequency and up to 2 times the expected frequency based on estimated disease prevalence
Very likely pathogenic	Report as "reported disease associated mutation"	N/A	• Variant convincingly segregates with disease in multiple families and is absent or very infrequent* in a large number of controls ---------- OR ---------- • Variant convincingly segregates with disease in one family, has been observed in multiple unrelated probands, and is absent or very infrequent* in controls

			---------- OR ---------- • Variant has been observed in multiple unrelated probands, and is absent or very infrequent* in controls -- AND – Available functional and/ or animal model evidence suggests evidence of functionality/disruptive effect on normal gene function
* As defined by an allele frequency consistent with mode of inheritance and population prevalence of disease.			

Table S2. Phenotypic characterization of inherited disease conditions.

Associated phenotype/condition	Definition
Inheritance	Dominant, Recessive, X-linked, Multifactorial
Onset- Earliest Onset- Median	Infantile, childhood, adolescence, early-adulthood, mid-adulthood, late-adulthood
Severity if untreated Severity with treatment	Rarely affects quality of life, mild impact on quality or quantity of life, moderate impact on quality or quantity of life, severely limits either quality or quantity of life, severely limits both quality and quantity of life
Actionability and Efficacy of action	Medical or surgical prevention available; cure available; condition is not curable, but effective protection from most decrements in quality and quantity of life; treatments available but only minimally affect decrements in quality and quantity of life; only palliative care available
Penetrance and expressivity	Near complete or complete, high, reduced or low penetrance If available, describe phenotypic variability in affected individuals For recessive conditions: do carriers manifest any clinical phenotype?
Prevalence	Overall and population specific, when described

Table S3. Common single nucleotide variants used for diabetes genetic risk assessment in Caucasians.

Chromosome	rsid	hg19 position	Alternative allele	Frequency	Odds ratio
1	rs10923931	120319482	T	0.06	1.10
1	rs340874	212225879	C	0.54	1.08
2	rs780094	27594741	C	0.60	1.04
2	rs11899863	43472323	C	0.92	1.15
2	rs243021	60438323	A	0.45	1.09
2	rs7593730	160879700	C	0.83	1.11
2	rs3923113	165210095	A	0.36	1.04
2	rs7578326	226728897	A	0.68	1.10
3	rs13081389	12264800	A	0.97	1.21
3	rs6795735	64680405	C	0.52	1.07
3	rs11708067	124548468	A	0.80	1.10
3	rs1470579	187011774	C	0.28	1.12
4	rs1801214	6353923	T	0.67	1.12
5	rs459193	55842508	G	0.75	1.05
6	rs10440833	20796100	A	0.25	1.22
7	rs2191349	15030834	T	0.56	1.07
7	rs849134	28162747	A	0.51	1.12
7	rs4607517	44202193	A	0.22	1.05
7	rs972283	130117394	G	0.54	1.10
8	rs516946	41638405	C	0.80	1.10
8	rs896854	96029687	T	0.49	1.09
8	rs3802177	118254206	G	0.72	1.16
9	rs944801	22041670	C	0.58	1.09
9	rs10965250	22123284	G	0.76	1.19
9	rs13292136	81141948	C	0.94	1.19
9	rs2796441	83498768	G	0.62	1.07
10	rs12779790	12368016	A	0.23	1.09
10	rs7903146	114748339	T	0.31	1.40

11	rs2334499	1653425	T	0.42	1.07
11	rs5215	17365206	C	0.43	1.08
11	rs1552224	72110746	A	0.87	1.13
11	rs1387153	92313476	T	0.23	1.13
12	rs10842994	27856417	C	0.83	1.09
12	rs1531343	64461161	C	0.10	1.15
12	rs4760915	69920379	T	0.25	1.09
12	rs4760790	69921061	A	0.26	1.10
12	rs7957197	119945069	T	0.85	1.12
15	rs7163757	60178900	C	0.54	1.06
15	rs7178572	75534245	G	0.68	1.08
15	rs11634397	78219277	G	0.62	1.09
15	rs8042680	89322341	A	0.24	1.07
16	rs11642841	52402988	A	0.46	1.14
16	rs7202877	73804746	T	0.91	1.15
17	rs4430796	33172153	G	0.49	1.13
18	rs12970134	56035730	A	0.28	1.08
19	rs10401969	19268718	C	0.09	1.13
20	rs4812829	42422681	A	0.20	1.07

Table S4. Common single nucleotide variants used for coronary artery disease genetic risk assessment in Caucasians.

Chromosome	rsid	hg19 position	Alternative allele	Frequency	Odds ratio
1	rs11206510	55268627	T	0.82	1.06
1	rs17114036	56735409	A	0.91	1.11
1	rs602633	109623034	C	0.78	1.12
1	rs4845625	152688691	T	0.44	1.05
1	rs17465637	220890152	C	0.72	1.05
2	rs515135	21139562	G	0.83	1.08
2	rs6544713	43927385	T	0.29	1.06
2	rs1561198	85663500	A	0.45	1.05
2	rs2252641	145517931	G	0.45	1.05
2	rs2351524	203589237	T	0.12	1.12
3	rs2306374	139602642	C	0.16	1.08
4	rs7692387	156854759	G	0.80	1.07
5	rs273909	131695252	C	0.13	1.08
6	rs12526453	13035530	C	0.68	1.09
6	rs12205331	35006433	C	0.79	1.04
6	rs10947789	39282900	T	0.76	1.06
6	rs12190287	134256218	C	0.61	1.07
6	rs2048327	160783522	G	0.37	1.06
6	rs4252120	161063598	T	0.72	1.06
7	rs2023938	19003300	G	0.11	1.08
7	rs11556924	129450732	C	0.64	1.09
8	rs264	19857460	G	0.85	1.07
8	rs2954029	126560154	A	0.54	1.05
9	rs3217992	21993223	A	0.38	1.16
9	rs1333049	22115503	C	0.47	1.23
9	rs495828	135144688	T	0.20	1.07
10	rs2505083	30375128	C	0.42	1.06
10	rs2047009	43859919	C	0.50	1.05

10	rs501120	44073873	A	0.84	1.07
10	rs2246833	90995834	T	0.37	1.06
11	rs974819	103165777	A	0.29	1.07
11	rs9326246	116116943	C	0.10	1.09
12	rs3184504	110368991	T	0.42	1.07
13	rs9319428	27871621	A	0.32	1.06
13	rs4773144	109758713	G	0.43	1.07
13	rs9515201	109838799	A	0.30	1.06
14	rs2895811	99203695	C	0.43	1.06
15	rs11072794	76793637	T	0.28	1.07
15	rs7173743	76928839	T	0.57	1.07
15	rs17514846	89217554	A	0.44	1.06
17	rs2281727	2064695	C	0.36	1.05
17	rs12936587	17484447	G	0.58	1.06
17	rs15563	44360192	C	0.53	1.04
19	rs1122608	11024601	G	0.76	1.10
19	rs2075650	50087459	G	0.14	1.11

Table S5. Common single nucleotide variants used for diabetes genetic risk assessment in East Asians.

Chromosome	rsid	hg19 position	Alternative allele	Frequency	Odds ratio
2	rs1260326	27584444	T	0.49	0.89
2	rs2943632	226762625	A	0.08	0.82
3	rs831571	64023337	T	0.39	0.92
3	rs6769511	187012984	C	0.26	1.17
4	rs6815464	1299901	G	0.42	0.88
6	rs9356744	20793465	C	0.36	1.14
6	rs9470794	38214822	C	0.27	1.12
6	rs1535500	39392028	T	0.42	1.08
7	rs6467136	126952194	A	0.21	0.9
8	rs11774700	118289451	C	0.45	0.9
9	rs12378556	4251496	G	0.38	1.11
9	rs10811661	22124094	C	0.41	0.85
10	rs10882091	94364357	C	0.08	1.3
11	rs2074314	17368397	G	0.37	1.15
15	rs3743451	89305185	T	0.48	0.9
16	rs17797882	77964419	T	0.32	1.08
16	rs16955379	80046874	T	0.2	0.93
19	rs3786897	38584848	G	0.44	0.91

Table S6. Common single nucleotide variants used for coronary artery disease genetic risk assessment in East Asians.

Chromosome	rsid	hg19 position	Alternative allele	Frequency	Odds ratio
4	rs1842896	156730909	T	0.76	1.14
6	rs9349379	13011943	G	0.74	1.15
6	rs9268402	32449331	G	0.59	1.16
6	rs12524865	134238367	C	0.61	1.11
9	rs10757274	22086055	G	0.46	1.37
9	rs1333042	22093813	G	0.67	1.37
12	rs7136259	88605319	T	0.39	1.11
12	rs11066280	111302166	A	0.17	1.19

Table S7. Healthcare Common Procedure Coding System (HCPCS) codes for proposed secondary clinical evaluations

Test or procedure	HCPCS code
Audiometry	92557
Spirometry with bronchodilator challenge	94060
Neutrophil oxidative burst	86352
vWF antigen	85246
vWF multimers	85247
Ristocetin cofactor	85245
Factor VIII activity	85240
Factor XIII activity	85290
Amylase, serum	82150
Lipase, serum	83690
Lipid panel, serum	80061
Renal ultrasound	76775
Transthoracic echocardiogram, full	93306
12 lead electrocardiogram	93000
Colonoscopy, high risk Medicare patient	G0105
High complexity office consultation	99245
Urinalysis, non-automated, with microscopy	81000
Urinalysis, automated, without microscopy	81003
Hemoglobin A1c, serum	83036
Glucose, serum	82947
Basic metabolic panel, serum	80048
Ferritin, serum	82728
Iron, serum	83540
Transferrin, serum	84466
Iron binding capacity, serum	83550
Complete blood count, serum	85025
Hepatic function panel, serum	80076
Alpha 1 antitrypsin, serum	82104
Prothrombin time, serum	85610
Partial thromboplastin time, serum	85730
Calcium, serum	82310
Phosphorus, serum	84100
Parathyroid hormone, serum	83970
Homocysteine, serum	83090
Thyroid stimulating hormone, serum	84443
Free T4, serum	84439
Cortisol, serum	82533
Insulin-like growth factor 1, serum	84305
Insulin-like growth factor binding protein 3, serum	82397

Table S8. Inter-rater agreement in pathogenicity classification of 18 randomly selected rare genetic variants

Variant \ Rater	1	2	3	4	5	6
1	1	6	--	--	--	7
2	2	3	--	--	--	1
3	4	6	--	--	--	7
4	6	7	--	--	--	7
5	1	0	--	--	--	1
6	1	0	--	--	--	1
7	2	--	3	--	--	--
8	5	--	5	--	--	--
9	1	--	0	--	--	--
10	0	--	6	--	--	--
11	7	--	7	--	--	--
12	6	--	6	--	--	--
13	1	--	--	1	1	--
14	5	--	--	1	5	--
15	3	--	--	1	1	--
16	3	--	--	1	1	--
17	5	--	--	5	5	--
18	3	--	--	3	3	--

Key: 0, likely benign, strong evidence; 1, some evidence suggesting benign; 2, novel, predicted benign; 3, uncertain significance; 4, novel, predicted damaging by 1 program; 5, novel, predicted damaging by 2 or more programs; 6, some evidence suggesting pathogenic; 7, likely pathogenic; 8, very likely pathogenic.

Table S9. Inter-rate agreement in suitability for reporting of 18 randomly selected rare genetic variants

Variant \ Rater	1	2	3	4	5	6
1	0	0	--	--	--	1
2	0	0	--	--	--	0
3	1	1	--	--	--	1
4	1	1	--	--	--	1
5	0	0	--	--	--	0
6	0	0	--	--	--	0
7	0	--	0	--	--	--
8	0	--	0	--	--	--
9	1	--	1	--	--	--
10	0	--	0	--	--	--
11	1	--	1	--	--	--
12	1	--	1	--	--	--
13	0	--	--	0	0	--
14	1	--	--	0	1	--
15	1	--	--	0	0	--
16	0	--	--	0	0	--
17	1	--	--	1	1	--
18	1	--	--	1	1	--

Key: 0, variant not suitable for reporting; 1, variant suitable for reporting.

Table S10. Genetic variants classified as "disease causing" or "likely disease causing" by HGMD in ACMG reportable genes.

rsid or hg19 coordinate and risk allele	Variant type	Gene	Disease association- ACMG	Disease association – HGMD	HGMD classification	MDGT variant classification	MDGT reporting classification	Ethnically-matched population allele frequency*	Frequency in study cohort	MDGT Classification comments
rs1042023	Missense SNV	*APOB*	Familial hyperlipidemia	Familial hypolipidemia	Likely disease causing mutation	VUS	Not reported	1%	1/12	Associated with hypolipidemia, not hyperlipidemia
rs80359876	Frameshift indel	*BRCA1*	HBOC	HBOC	Disease causing mutation	Very likely pathogenic	Previously reported disease associated variant	--	1/12	Has been observed in multiple unrelated probands with HBOC; absent in controls; functional evidence to support pathogenicity; personal communication with private sequencing provider suggests variant is pathogenic
rs61757642	Splice disrupting SNV	*BRCA2*	HBOC	HBOC	Disease causing mutation	VUS	VUS	0.03%	1/12	Conflicting evidence of pathogenicity in literature; personal communication with private sequencing provider suggests variant may be benign
rs1801426	Missense SNV	*BRCA2*	HBOC	HBOC	Likely disease causing mutation	Some evidence suggesting	Not reported	10%	1/12	Common polymorphism

						benign				
rs79483201	Missense SNV	*BRCA2*	HBOC	HBOC	Disease causing mutation	VUS	VUS	1%	1/12	Polymorphism; conflicting evidence of pathogenicity in literature; personal communication with private sequencing provider suggests variant may be benign
chr18: 29116333 T>G	Missense SNV	*DSG2*	Arrhythmogenic right ventricular cardiomyopathy	Arrhythmogenic right ventricular cardiomyopathy	Disease causing mutation	Some evidence suggesting pathogenic	VUS	--	1/12	
rs12720449	Missense SNV	*KCNQ1*	Long QT syndrome	Long QT syndrome	Likely disease causing mutation	Some evidence suggesting benign	Not reported	11%	1/12	Common polymorphsim
rs2228671	Nonsense SNV	*LDLR*	Familial hyperlipidemia	Hypercholesterolemia	Likely disease causing mutation	Some evidence suggesting benign	Not reported	13%	1/12	Common polymorphsim associated with complex hyperlipidemia
rs5926	Synonymous SNV	*LDLR*	Familial hyperlipidemia	Hypercholesterolemia	Likely disease causing mutation	Some evidence suggesting benign	Not reported	1%	1/12	Polymorphism associated with complex hyperlipidemia
chr2:47705625 G>A	Missense SNV	*MSH2*	Lynch syndrome	Muir-Torre syndrome	Disease causing mutation	VUS	VUS	0.17%	1/12	Described disease associated mutation is an indel at the same genomic position as this missense variant
rs3729989	Missense SNV	*MYBPC3*	Familial hypertrophic cardiomyopathy	Familial hypertrophic cardiomyopathy	Likely disease causing mutation	Some evidence suggesting benign	Not reported	4%/13%	2/12	Common polymorphism
rs10254120	Missense SNV	*PMS2*	Lynch syndrome	Hereditary non-polyposis colon	Likely disease causing	Some evidence	Not reported	9%	1/12	Common polymorphism

				cancer	mutation	suggesting benign				
chr12:33049482 G>T	Missense SNV	*PKP2*	Arrhythmogenic right ventricular cardiomyopathy	Arrhythmogenic right ventricular cardiomyopathy	Disease causing mutation	VUS	VUS	0.04%	1/12	Lack of clear segregation with disease in published reports
rs35118262	Missense SNV	*RET*	MEN2, familial medullary thyroid cancer	Hirschsprung disease	Disease causing mutation	VUS	Not reported	1%	1/12	Polymorphism, described association is with Hirschsprung disease
rs45517323	Missense SNV	*TSC2*	Tuberous sclerosis	Tuberous sclerosis	Likely disease causing mutation	Some evidence suggesting benign	Not reported	3%	1/12	Polymorphism

* Highest reported population allele frequency from an ethnically---matched genetic survey (HapMap2 and 3, 1000 genomes phase 1 or pilot 1, NHLBI Grand Opportunity exome sequencing project, Complete Genomics publicly available genomes).

Abbreviations: ACMG, American College of Genetics and Genomics; HBOC, hereditary breast and ovarian cancer; HGMD, Human Gene Mutation Database; MDGT, multi-disciplinary genomics team; MEN2, multiple endocrine neoplasia type 2; SNV, single nucleotide variant; VUS, variant of uncertain significance.

Table S11. Population prevalence of genetic variants classified as "disease causing" by HGMD and loss-of-function variants in ACMG incidental finding genes in 1,092 subjects sequenced in the 1000 genomes study, phase 1.

Gene	Subjects with one or more damaging mutations of any frequency, n (%)		Subjects with one or more damaging mutations with allele frequency < 5%, n (%)		Subjects with one or more damaging mutations with allele frequency < 0.5%, n (%)	
	HGMD - DM	LOF	HGMD - DM	LOF	HGMD -DM	LOF
All genes	1091 (99.9)	433 (39.7)	657 (60.2)	153 (14.01)	373 (34.2)	37 (3.4)
BRCA2	1091 (99.9)	13 (1.2)	61 (5.6)	13 (1.2)	46 (4.2)	1 (0.1)
*RYR1**	139 (12.7)	--	63 (5.8)	--	38 (3.5)	--
BRCA1	33 (3.0)	1 (0.1)	33 (3.0)	1 (0.1)	33 (3.0)	1 (0.1)
MYBPC3	45 (4.1)	1 (0.1)	45 (4.1)	1 (0.1)	30 (2.8)	1 (0.1)
SCN5A	80 (7.3)	1 (0.1)	80 (7.3)	1 (0.1)	29 (2.7)	1 (0.1)
PKP2	26 (2.4)	6 (0.6)	26 (2.4)	6 (0.6)	26 (2.4)	6 (0.6)
LDLR	24 (2.2)	2 (0.2)	24 (2.2)	2 (0.2)	24 (2.2)	2 (0.2)
*RET**	24 (2.2)	--	24 (2.2)	--	24 (2.2)	--
APC	45 (4.1)	0 (0)	45 (4.1)	0 (0)	20 (1.8)	0 (0)
MLH1	18 (1.7)	1 (0.1)	18 (1.7)	1 (0.1)	18 (1.7)	1 (0.1)
MSH6	18 (1.7)	19 (1.7)	18 (1.7)	19 (1.7)	18 (1.7)	1 (0.1)
TSC1	59 (5.4)	0 (0)	59 (5.4)	0 (0)	15 (1.4)	0 (0)
DSG2	94 (8.6)	0 (0)	94 (8.6)	0 (0)	14 (1.3)	0 (0)
DSP	46 (4.2)	2 (0.2)	46 (4.2)	2 (0.2)	14 (1.3)	2 (0.2)
*MYH7**	23 (2.1)	--	23 (2.1)	--	12 (1.1)	--
RB1	12 (1.1)	0 (0)	12 (1.1)	0 (0)	12 (1.1)	0 (0)
*RYR2**	12 (1.1)	--	12 (1.1)	--	12 (1.1)	--
MSH2	28 (2.6)	0 (0)	28 (2.6)	0 (0)	11 (1.0)	0 (0)
FBN1	47 (4.3)	0 (0)	47 (4.3)	0 (0)	9 (0.8)	0 (0)
KCNH2	8 (0.7)	1 (0.1)	8 (0.7)	1 (0.1)	8 (0.7)	1 (0.1)
MUTYH	8 (0.7)	13 (1.2)	8 (0.7)	13 (1.2)	8 (0.7)	13 (1.2)
*PCSK9**	23 (2.1)	--	23 (2.1)	--	8 (0.7)	--
*APOB**	649 (59.4)	--	116 (10.6)	--	6 (0.6)	--
KCNQ1	6 (0.6)	0 (0)	6 (0.6)	0 (0)	6 (0.6)	0 (0)
TP53	6 (0.6)	1 (0.1)	6 (0.6)	1 (0.1)	6 (0.6)	1 (0.1)
TSC2	6 (0.6)	0 (0)	6 (0.6)	0 (0)	6 (0.6)	0 (0)
WT1	4 (0.4)	0 (0)	4 (0.4)	0 (0)	4 (0.4)	0 (0)
*CACNA1S**	3 (0.3)	--	3 (0.3)	--	3 (0.3)	--
SDHB	22 (2.0)	1 (0.1)	22 (2.0)	1 (0.1)	3 (0.3)	1 (0.1)
DSC2	2 (0.2)	13 (1.2)	2 (0.2)	13 (1.2)	2 (0.2)	1 (0.1)
COL3A1	14 (1.3)	79 (7.2)	14 (1.3)	79 (7.2)	1 (0.1)	1 (0.1)
LMNA	1 (0.1)	339 (31.0)	1 (0.1)	0 (0)	1 (0.1)	0 (0)
MEN1	1 (0.1)	0 (0)	1 (0.1)	0 (0)	1 (0.1)	0 (0)
NF2	1 (0.1)	2 (0.2)	1 (0.1)	2 (0.2)	1 (0.1)	2 (0.2)
TGFBR2	1 (0.1)	1 (0.1)	1 (0.1)	1 (0.1)	1 (0.1)	1 (0.1)
ACTA2	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
*ACTC1**	0 (0)	--	0 (0)	--	0 (0)	--

GLA	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
MYH11	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
*MYL2**	0 (0)	--	0 (0)	--	0 (0)	--
*MYL3**	0 (0)	--	0 (0)	--	0 (0)	--
MYLK	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
PMS2	149 (13.6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
*PRKAG2**	0 (0)	--	0 (0)	--	0 (0)	--
PTEN	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
*SDHAF2**	0 (0)	--	0 (0)	--	0 (0)	--
SDHC	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
SDHD	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
SMAD3	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
STK11	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
TGFBR1	0 (0)	1 (0.1)	0 (0)	1 (0.1)	0 (0)	1 (0.1)
TMEM43	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
*TNNI3**	0 (0)	--	0 (0)	--	0 (0)	--
TNNT2	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
*TPM1**	0 (0)	--	0 (0)	--	0 (0)	--
VHL	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

*Gene not subject to reporting loss of function variants per ACMG incidental findings recommendations.

Abbreviations: ACMG, American College of Medical Genetics and Genomics; LOF, loss-of-function, defined as splice disrupting, frameshift indel, and stopgain or stoploss; HGMD, Human Gene Mutation Database; HGMD---DM, variants classified as “disease causing” by the Human Gene Mutation Database.

Table S12. Pharmacogenomic findings with Clinical Pharmacogenomic Implementation Guideline support for change in drug dose or administration in whole genome sequence data.

Diplotype-based metabolic phenotype			
Clopidogrel – CYP2C19	**Implication for drug**	**Recommendation for drug use**	**N (total n = 12)**
Ultra-rapid metabolizer (e.g. *1/*17)	Increased platelet inhibition	Label recommended dosage and administration	0
Extensive metabolizer (*1/*1)	Normal platelet inhibition	Label recommended dosage and administration	5
Intermediate metabolizer (e.g. * 1/*2)	Reduced platelet inhibition	Alternative antiplatelet therapy	7
Poor metabolizer (e.g. *2/*2)	Significantly reduced platelet inhibition	Alternative antiplatelet therapy	0
Codeine – CYP2D6			
Ultra-rapid metabolizer (e.g. *1/*2xN)	Increased formation of morphine	Avoid codeine use due to potential for toxicity	0
Extensive metabolizer (e.g. *1/*1)	Normal morphine formation	15-50 mg every 4h as needed	11
Intermediate metabolizer (e.g. *4/*10)	Reduced morphine formation	Begin with 15-60 mg every 4h as needed	0
Poor metabolizer (e.g. *4/*4)	Greatly reduced morphine formation	Avoid codeine use due to lack of efficacy	1
Simvastatin - rs4149056			
Normal activity (TT)	Low myopathy risk	Prescribe desired starting dose.	11
Intermediate activity (TC)	Intermediate myopathy risk	Consider a lower dose; if suboptimal efficacy, consider an alternative statin.	1
Low activity (CC)	High myopathy risk	Prescribe a lower dose or consider an alternative statin	0
Thiopurines – TPMT			
High activity (e.g. *1/*1)	Lower concentrations of TGN metabolites, higher methylTIMP	Start with normal starting dose and adjust doses of thiopurine and of other myelosuppressive therapy without any special emphasis on thiopurine. Allow 2 weeks to reach steady state after each dose adjustment.	12
Intermediate activity (e.g. *1/*2)	Moderate to high concentrations of TGN metabolites; low concentrations of methylTIMP	Start with reduced doses and adjust doses based on degree of myelosuppression and disease-specific guidelines. Allow 2-4 weeks to reach steady state after each dose adjustment	0
Low or deficient activity	Extremely high	For malignancy, start with	0

(e.g. *3A/*3A)	concentrations of TGN metabolites; no methylTIMP metabolites	drastically reduced doses and adjust doses based on degree of myelosuppression and disease-specific guidelines. Allow 4-6 weeks to reach steady state after each dose adjustment.	
Warfarin – CYP2C9; VKORC1 rs9923231*			
Normal activity (e.g. *1/*1;GG)	Normal dose requirement	5-7 mg/day	4
Intermediate activity (*1/*2;GA)	Intermediate dose requirement	3-4 mg/day	8
Low activity (*2/*2;AA)	Low dose requirement	0.5-2 mg/day	0
Any diplotype affecting drug dose or administration	--	--	11
* Based on United States Food and Drug Administration product insert recommendations. Abbreviations: TGN, thioguanine; TIMP - thiomercaptopurine. “Diplotype” refers to the specific combination of two haplotypes.			

Table S13. Burden and cost of proposed clinical follow-up

	Medical geneticist 1		Medical geneticist 2		Primary care physician 1		Primary care physician 2		Primary care physician 3	
Type	**n**	**Cost, USD***	**n**	**Cost, USD***	**n**	**Cost, USD***	**n**	**Cost, USD***	**n**	**Cost, USD***
Laboratory tests	0 (0-2)	0 (0-22)	0 (0-7)	0 (0-211)	2 (1-7)	32 (7-237)	0 (0-1)	0 (0-15)	2 (1-4)	27 (22-50)
Noninvasive diagnostic tests	0 (0-3)	0 (0-291)	1 (0-2)	18 (0-208)	0 (0-2)	0 (0-190)	0 (0-1)	0 (0-18)	0 (0-2)	0 (0-208)
Invasive diagnostic tests	0 (0-0)	0 (0-0)	0 (0-1)	0 (0-411)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-0)	0 (0-1)	0 (0-411)
Referrals	1 (1-3)	216 (216-649)	0 (0-2)	216 (0-433)	1 (0-3)	325 (0-649)	0 (0-2)	0 (0-433)	1 (0-2)	216 (0-433)
Total secondary clinical evaluations	3 (1-7)	286 (216-662)	3 (0-10)	433 (0-645)	3 (2-10)	436 (23-753)	1 (0-2)	11 (0-433)	3 (1-9)	242 (22-1087)
Total costs†	--	626 (557-1085)	--	773 (340-986)	--	776 (563-1093	--	351 (340-773)	--	582 (362-1427)

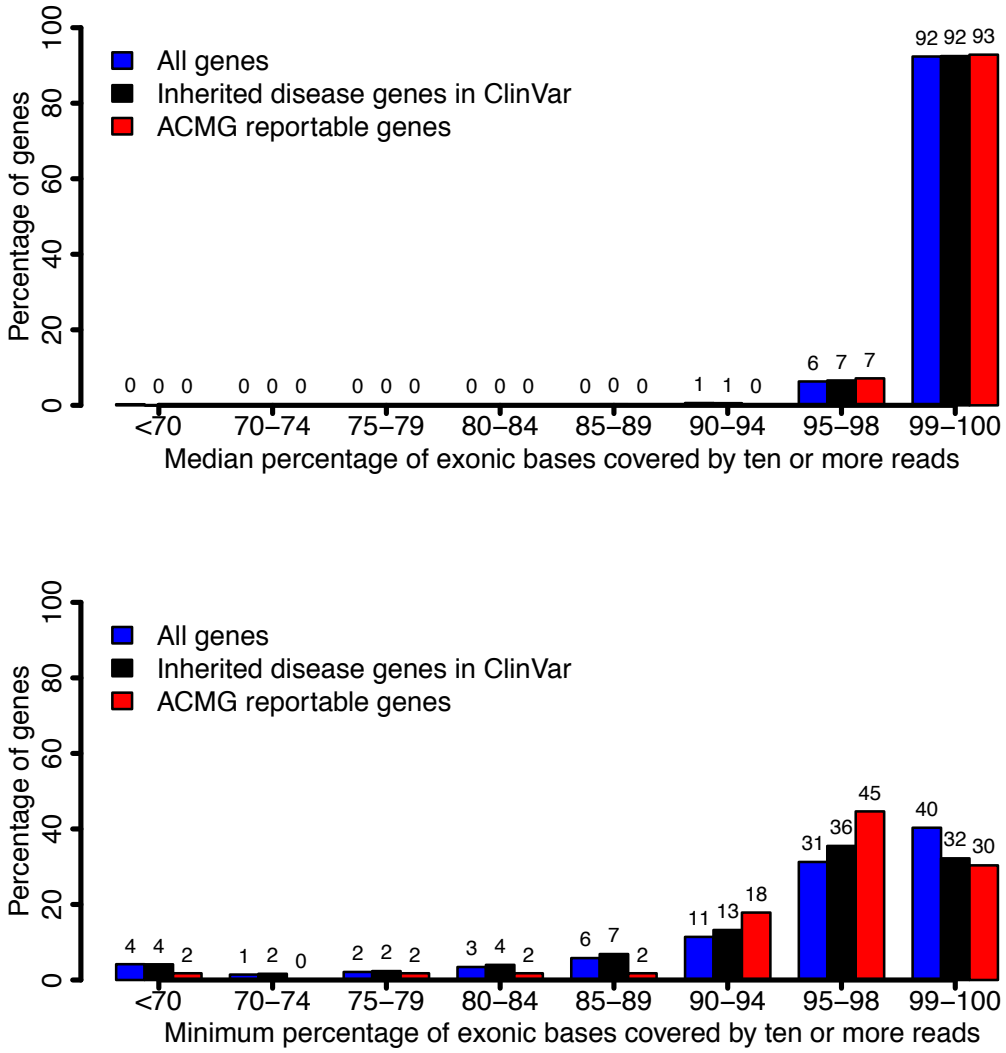
All values expressed as median (range) per individual.

*Estimated according to Centers for Medicare and Medicaid Services calendar year 2013 fee schedule for non-facility billing and Clinical Laboratory Fee national mid-point.

†Total cost estimates include a high complexity established patient visit and 120 minute genetic counseling session (HCPCS codes 99215 and 96040 (x4), respectively) for all participants.
Abbreviations: USD, United States dollars.

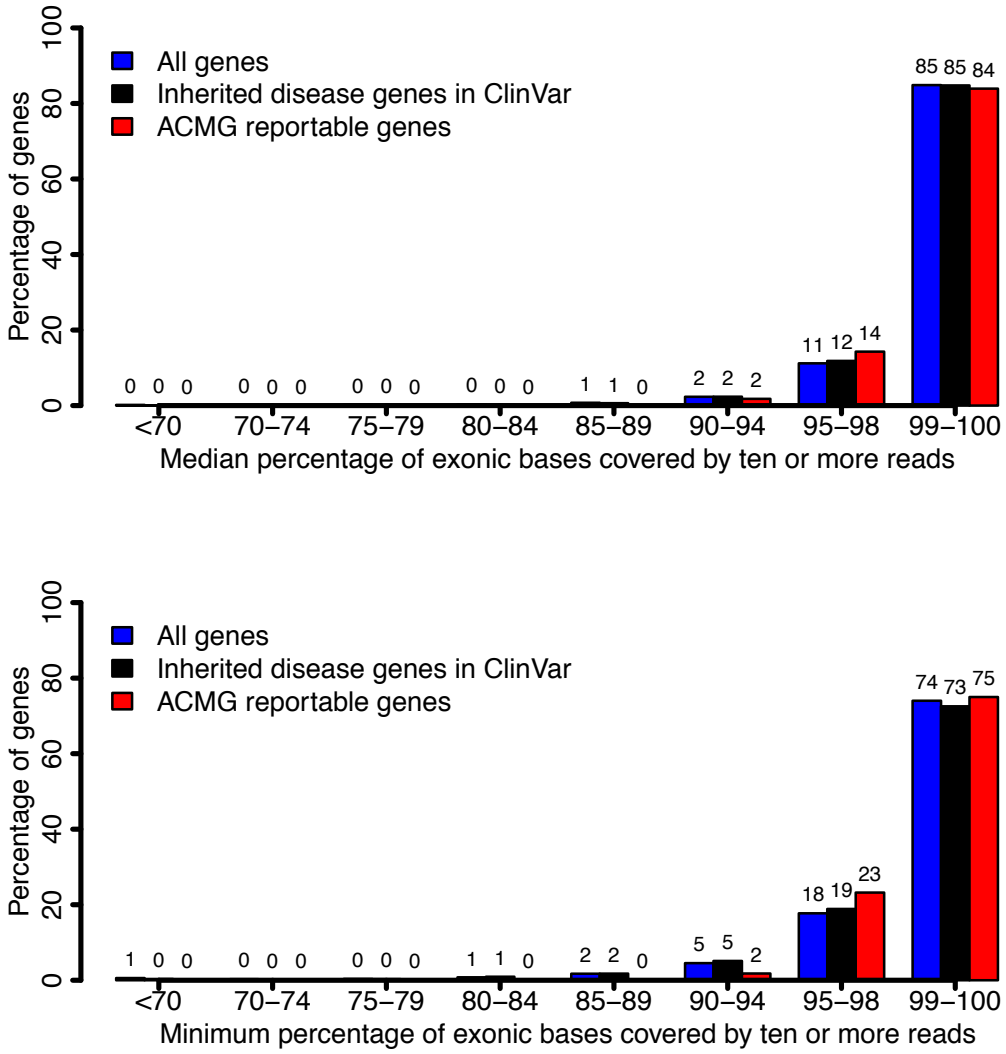
Supplementary Figures

Figure S1. Exonic sequence coverage by the Illumina platform.



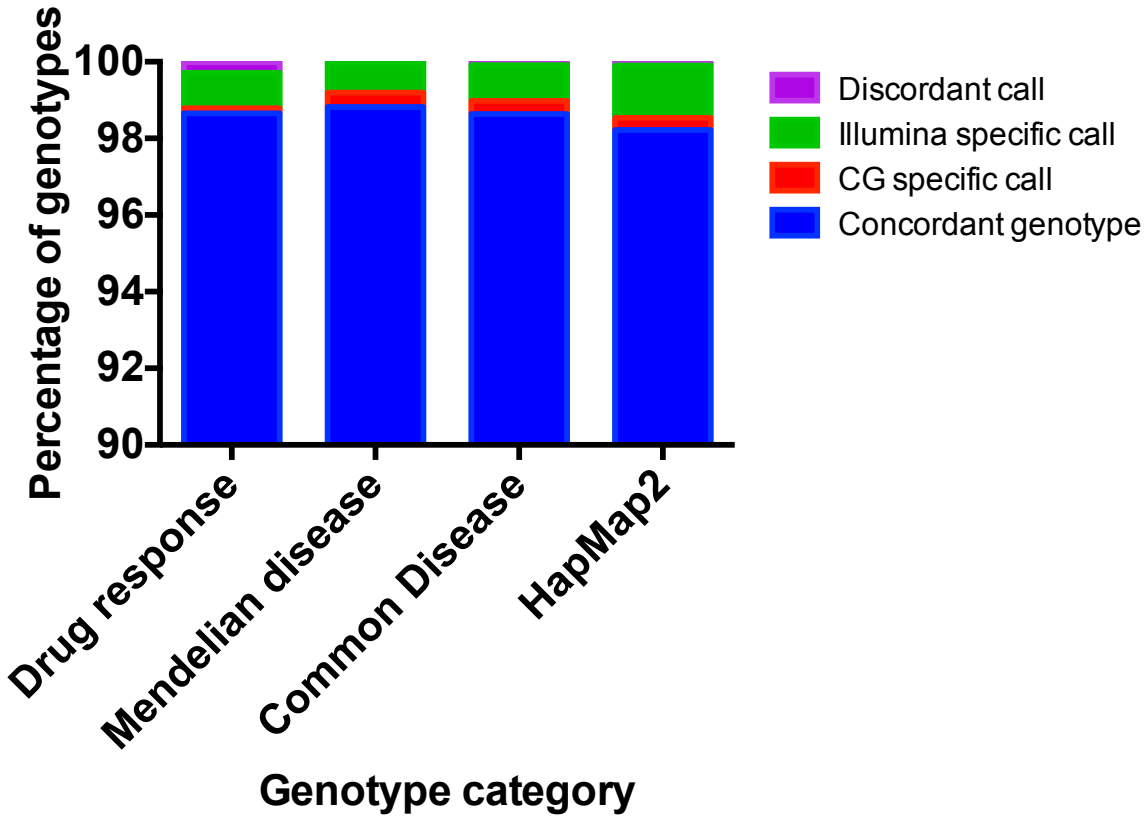
Top, median percentage of exonic bases covered by ten or more reads. Bottom, minimum percentage of exonic bases covered by ten or more reads among the ten sequenced participants. The minimum percentage of genes covered defines the percentage of genes always covered above specific thresholds for sequence depth and percentage of exonic bases, i.e., the percentage of genes covered in all ten subjects at or above such thresholds. Coverage calculations were performed by the "CGA" tool provided by Complete Genomics, Inc. for all transcripts corresponding to all human Entrez IDs ("All genes"), Entrez IDs representing 2,716 genes cataloged in ClinVar in association with inherited disease genes ("inherited disease genes in ClinVar"), and Entrez IDs corresponding to 56 genes recommended by the American College of Medical Genetics and Genomics for discovery and return of secondary findings in WGS/WES.

Figure S2. Exonic sequence coverage by the Complete Genomics platform.



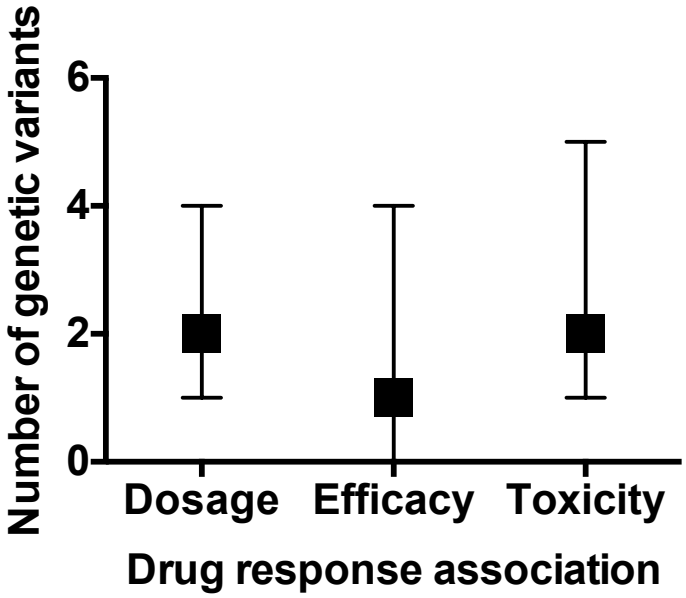
Top, median percentage of exonic bases covered by ten or more reads. Bottom, minimum percentage of exonic bases covered by ten or more reads among the ten sequenced participants. The minimum percentage of genes covered defines the percentage of genes always covered above specific thresholds for sequence depth and percentage of exonic bases, i.e., the percentage of genes covered in all ten subjects at or above such thresholds. Coverage calculations were performed by the "CGA" tool provided by Complete Genomics, Inc. for all transcripts corresponding to all human Entrez IDs ("All genes"), Entrez IDs representing 2,716 genes cataloged in ClinVar in association with inherited disease genes ("Inherited disease genes in ClinVar"), and Entrez IDs corresponding to 56 genes recommended by the American College of Medical Genetics and Genomics for discovery and return of secondary findings in WGS/WES.

Figure S3. Cross-platform genotype concordance at previously reported variant loci.



Abbreviations: CG, Complete Genomics. Drug response variants were defined as 555 variants cataloged by the Pharmacogenomics Knowledge Base as associated with clinical drug response. Inherited disease variants were defined as 72,383 (males) and 72,316 (females) variants cataloged in the Human Gene Mutation Database (HGMD) that reside in high impact disease genes cataloged by Clinvar, Common disease genotypes were defined as variants cataloged by the National Center for Biotechnology Information Genome Wide Association Study Catalog, database download 6.12.12.

Figure S4. Replicated pharmacogenomic findings in study subjects by association type.



Symbols represent sample median and error bars represent sample range. All genetic variants had PharmGKB level of evidence 1B or higher support for drug response associations.

References

1. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53-9.
2. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 2010;327:78-81.
3. Lam HY, Pan C, Clark MJ, et al. Detecting and annotating genetic variations using the HugeSeq pipeline. Nat Biotechnol 2012;30:226-9.
4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
5. Dewey FE, Chen R, Cordero SP, et al. Phased whole genome genetic risk in a family quartet using a major allele reference sequence. PLoS Genet 2011;7:e1002280.
6. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.
7. A haplotype map of the human genome. Nature 2005;437:1299-320.
8. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.
9. Mills RE, Pittard WS, Mullaney JM, et al. Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res 2011;21:830-9.
10. Lam HY, Mu XJ, Stutz AM, et al. Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. Nature Biotechnol 2010;28:47-55.
11. Chen K, Wallis JW, McLellan MD, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 2009;6:677-81.
12. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: An approach to discover, genotype and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res 2011.
13. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009;25:2865-71.
14. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 2012;28:i333-i9.
15. Miller CA, Hampton O, Coarfa C, Milosavljevic A. ReadDepth: a parallel R package for detecting copy number alterations from short sequencing reads. PloS One 2011;6:e16327.
16. Cooper DN, Ball EV, Krawczak M. The human gene mutation database. Nucleic Acids Res 1998;26:285-7.
17. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: 2008 update. Genome Med 2009;1:13.
18. Riggs ER, Wain KE, Riethmaier D, et al. Towards a Universal Clinical Genomics Database: The 2012 International Standards for Cytogenomic Arrays (ISCA) Consortium Meeting. Human Mut 2013.
19. Deloukas P, Kanoni S, Willenborg C, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013;45:25-33.

20. Lu X, Wang L, Chen S, et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. Nat Genet 2012;44:890-4.
21. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Gent 2012;44:981-90.
22. Cho YS, Chen CH, Hu C, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet 2012;44:67-72.
23. Whirl-Carrillo M, McDonagh EM, Hebert JM, et al. Pharmacogenomics knowledge for personalized medicine. Clinical Pharmacol Ther 2012;92:414-7.
24. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
25. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2011;12:745-55.
26. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genet Med 2008;10:294-300.
27. Knowles JW, Assimes TL, Kiernan M, et al. Randomized trial of personal genomics for preventive cardiology: design and challenges. Circ Cardiovasc Genet 2012;5:368-76.
28. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol 1989;129:687-702.

STANFORD GENOMIC MEDICINE APPLICATION PILOT PROJECT (GMAPP)

Personal genome sequence results for:

[redacted]

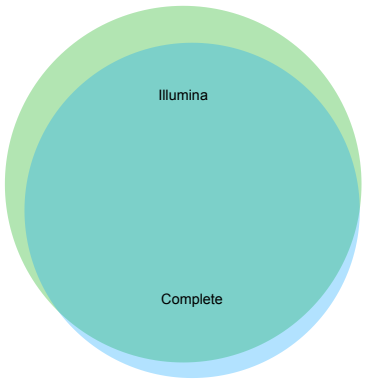
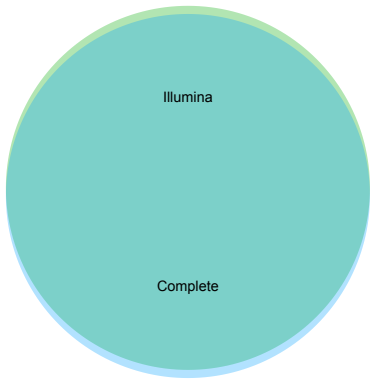
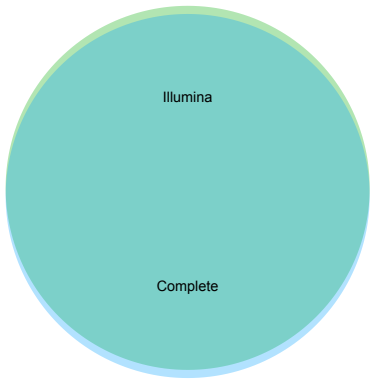
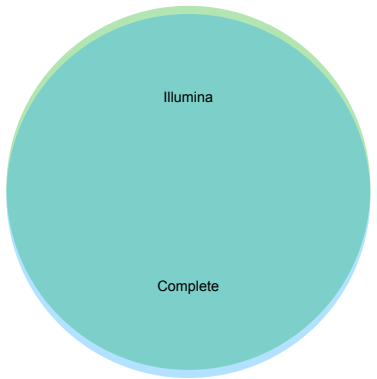
Whole genome sequence and analysis summary

Technical summary

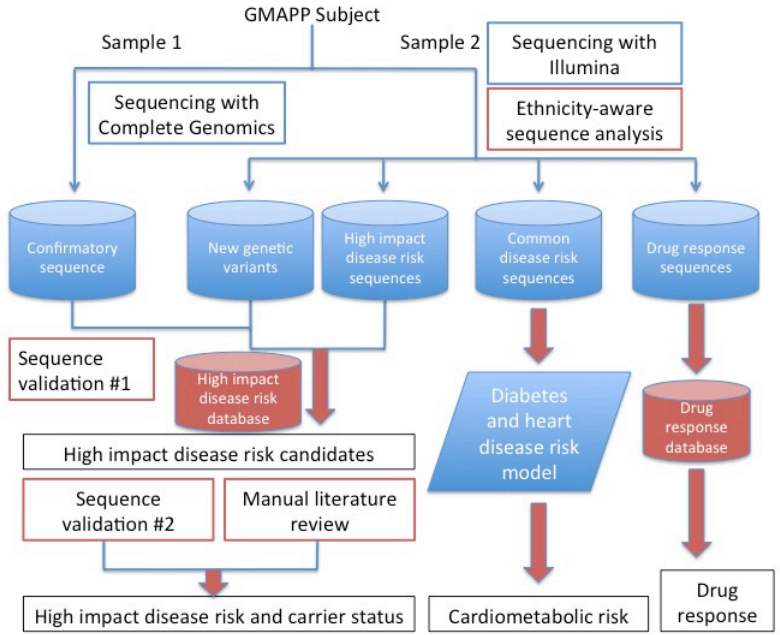
Whole genome sequence data was generated using paired end 101 base pair sequence by Illumina, Inc., and 36 base pair sequencing by Complete Genomics, Inc. Sequence reads were aligned to a reference genome representing the most common genetic code in Caucasian individuals. New and previously described genetic variation affecting rare, Mendelian single gene or "Mendelian" disorders, common complex diseases, and drug response was identified using sequence data from both platforms. In general, whole genome sequencing provides the most comprehensive map of personal genetic sequence. However, it is important to remember that *whole genome sequencing cannot provide information about some known genetic risk factors for disease or drug response, including many large genetic rearrangements, and cannot provide information about clinical, environmental, and undiscovered genetic risk factors for disease and drug response. This report is meant for research purposes only and is not intended or approved for clinical use.*

Sequencing platform	Bases sequenced (in billions)	% of callable genome covered	Average depth of coverage	Mendelian disease risk coverage	Common disease risk coverage	Drug response coverage
Illumina, Inc.	124.62	95.1%	42x	99.9%	99.2%	98.7%
Complete Genomics, Inc.	189.96	96.9%	61x	99.2%	98.7%	98.8%

Genotype concordance

All variants from reference	Mendelian disease risk variants	Common disease risk variants	Variants associated with drug response
			

Analysis summary



Variant type	Number
All single variants from an ethnically-concordant reference genome	2,406,485
All small insertions and deletions	566,475
All large rearrangements	2,249
Variants that have never been seen before	80,497
Variants occurring in protein coding regions	15,240
Variants changing protein sequence	5,964
Variants knocking out a gene copy	70

Medical genome sequence data was analyzed according to allele frequency in 8 population groups, predicted pathogenicity, disease association, if known, and evolutionary evidence of importance by sequence comparison with 46 lower animals. Data sources used for genome analysis are as follows (full list in **Appendix A**):

Databases used for allele frequency estimation: 25
Databases used for gene effect prediction: 3
Databases used for functional effect prediction: 7
Databases used for sequence comparison to lower animals: 5
Databases used for medical effect prediction: 7

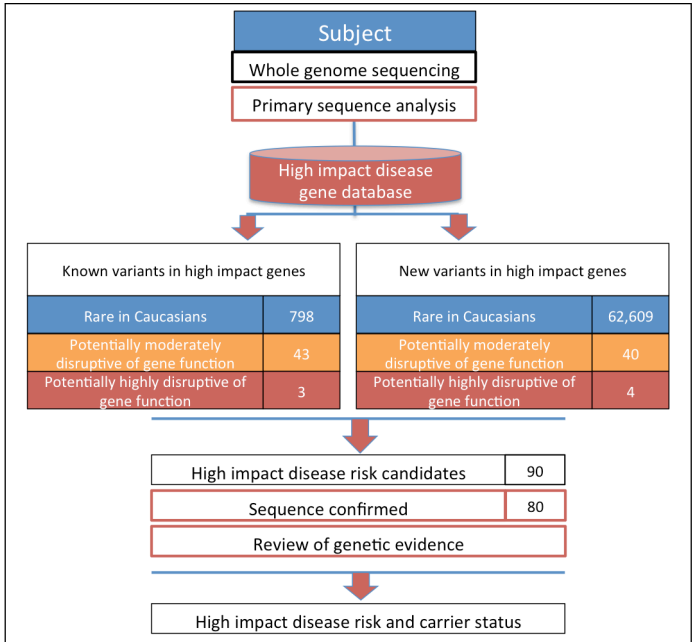
Mendelian disease risk and carrier status: Technical summary

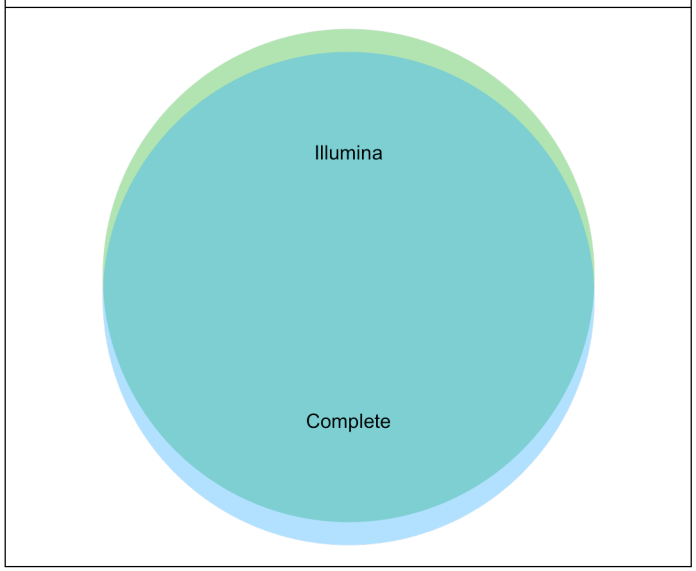
We scanned the genome for new and previously described genetic changes in 2,725 genes associated with 3,404 Mendelian disorders attributed to a single genetic variant, commonly referred to as "Mendelian disorders". In contrast to common diseases such as coronary artery disease and diabetes, these disorders are typically only found in small numbers of people but the associated individual genetic changes greatly increase the risk for disease.

Mendelian genetic diseases may require either one ("dominant") or two ("recessive") copies of a genetic risk variant to produce disease. We refer to individuals who carry one copy of a genetic variant causing recessive disease as "carriers." Generally, the main health implication for carriers is the risk of developing disease in future generations, though in some cases carriers may display mild manifestations of disease. First-degree relatives of carriers each have a 50% chance of carrying the disease-associated variant; if two carriers for the same disease condition have children, each child has a 25% chance of developing disease. First-degree relatives of individuals with genetic risk for dominant disease each have a 50% chance of having the same genetic risk. It is important to remember that *other unmeasured or undiscovered genetic and environmental factors may also influence risk for Mendelian disease.*

Of the 72,383 genetic variants that have been previously described in genes associated with Mendelian disorders, we were able to confidently sequence 72,284 (99.9%). Of these sequences, 798 differed from the most common sequence in Caucasian populations, 3 were highly disruptive to gene sequences, and 43 sequences were moderately disruptive of gene sequence or were very rare. We also observed 4 new highly disruptive genetic variants and 40 new moderately disruptive genetic variants in genes associated with Mendelian disorders.

Of these 90 candidate genetic variants, 80 (89.9%) were confirmed by the second sequencing platform. Genetic evidence supporting disease associations was manually reviewed for all variants by a board-certified genetic counselor and physician. A summary of our current knowledge about confirmed variants that may impact personal disease risk and carrier status is described below. **Appendix B** describes our criteria for reporting variants with potential Mendelian disease or carrier status implications.





Summary of Mendelian disease findings: Personal disease risk

Variant type	Variants
Personal disease risk	
Previously reported disease associated mutations	1
Rare, expected pathogenic variants with potential personal disease risk implications	0
Variants of uncertain significance	3

High level summary – Personal disease risk

Based on our current knowledge and analysis, we found one previously reported disease-associated mutation conferring susceptibility to acquired aplastic anemia. We found no rare, expected pathogenic variants with implications for personal Mendelian disease risk. We found three rare variants with uncertain personal disease risk implications in Mendelian disease-associated genes. Current genetic evidence is insufficient to confidently classify these variants as either disease associated or benign. Correlation with clinical and family history is advised.

Personal genetic risk for Mendelian disease

Previously reported disease associated mutations

Genetic Variant	Implication for personal disease risk	Disease description
TERT missense variant NM_198253: c.604G>A (p.A202T) g.chr5:1294397C>T In one copy of the *TERT* gene there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 0.016% of subjects sequenced as part of the NHLBI exome sequencing project.	This specific genetic variant has been described in several unrelated individuals with acquired aplastic anemia and is seen in low frequency in apparently unaffected populations.[1,2]	Aplastic anemia, also called bone marrow failure, occurs when the body does not make enough red blood cells, white blood cells and platelets. Several individuals with this particular genetic variant have developed acquired aplastic anemia. Mutations in the *TERT* gene also are the cause of a rare genetic syndrome called Dyskeratosis congenita (DC), which is an autosomal dominant or autosomal recessive telomere biology disorder, characterized by a classic triad of dysplastic nails, lacy reticular pigmentation of the upper chest and/or neck, and oral leukoplakia (patches on the tongue). People with DC are at increased risk for progressive bone marrow failure, myelodysplastic syndrome or acute myelogenous leukemia, solid tumors (usually squamous cell carcinoma of the head/neck or anogenital cancer), and pulmonary fibrosis. Individuals with one or more improperly functioning copies of this gene may be susceptible to acquired aplastic anemia. The prevalence of aplastic anemia is estimated to be 0.6-6.1 cases per million. Many individuals with genetic risk for acquired aplastic anemia do not develop clinical manifestations of the disease.

[1] Yamaguchi et al. N Engl J Med. 2005; 352:1413-24.
[2] Vulliamy et al. Blood Cells Mol Dis. 2005; 34:257-63.

Rare, expected variants with potential personal disease risk implications

We did not find any rare, expected pathogenic variants with personal disease risk implications.

Genetic variants of uncertain personal disease risk significance

Genetic Variant	Implication for personal disease risk	Disease description
EYA1 missense variant NM_172059: c. 1208A>T (p.Y403F) g.chr8:72128974T>A In one copy of the *EYA1* gene there is a sequence variant that changes the amino acid sequence of the gene. This variant has not been seen in any dataset so far and appears to be unique.	This novel variant is of uncertain significance with respect to personal risk of branciootorenal spectrum disorders, based on evidence of sequence conservation, prediction of biochemical effect, and rarity. Though this particular mutation has not been described in association with branciootorenal spectrum disorders, variants of this type have been implicated in the disorder.	Branchiootorenal syndrome (BOR) is characterized by malformations of the outer, middle, and inner ear associated with conductive, sensorineural, or mixed hearing impairment, branchial fistulae and cysts, and renal malformations ranging from mild renal hypoplasia to bilateral renal agenesis. Some individuals progress to end-stage renal disease in adulthood. *EYA1*-related BOR syndrome is inherited in an autosomal dominant manner, meaning that one or more copies of the *EYA1* gene must not function properly to cause the condition. The prevalence of branchiootorenal spectrum disorders is unknown. Most individuals with genetic risk do have clinical manifestations of the disease, and expressivity is highly variable.
COL1A1 missense variant NM_000088: c. 4181A>G (p.N1394S) g.chr17:48263206T>C In one copy of the *COL1A1* gene there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 0.26% of European subjects sequenced as part of the NHLBI exome sequencing project, and 0.4% of European subjects sequenced as part of the 1000Genomes study.	This rare variant is of uncertain significance with respect to personal risk of *COL1A1*-related osteogenesis imperfecta based on evidence of sequence conservation, prediction of biochemical effect, and rarity. Though this particular mutation has not been described in association with *COL1A1*-related OI, variants of this type have been implicated in this condition. This variant is not located within the domain typically implicated in disease, the collagen triple helical domain. The significance of variants outside the triple helical domain in this gene is not known. This variant is also of uncertain significance with respect to personal risk of Ehlers Danlos syndrome based on evidence of sequence conservation, prediction of biochemical effect, and rarity. One variant of this type has been identified in several individuals with a condition resembling classic Ehlers Danlos syndrome, though it was located within the typical collagen triple helical domain. The significance of variants outside the triple helical domain in this gene is not known.	*COL1A1*-related osteogenesis imperfecta (OI) is characterized by fractures with minimal or absent trauma, variable dentinogenesis imperfecta, and, in adult years, hearing loss. The clinical features of *COL1A1/2*-related OI represent a continuum ranging from perinatal lethality to individuals with severe skeletal deformities, mobility impairments, and very short stature to nearly asymptomatic individuals with a mild predisposition to fractures, normal dentition, normal stature, and normal life span. Classic EDS is a connective tissue disorder characterized by skin hyperextensibility, abnormal wound healing, and joint hypermobility. *COL1A1*-related OI and EDS are inherited in an autosomal dominant manner, meaning that one or more copies of the *COL1A1* gene must not function properly to cause the condition. The prevalence of COL1A1-related OI is unknown. OI of all types has a prevalence of approximately 6-7:100,000. Most individuals with genetic risk do have clinical manifestations of the disease, and expressivity is highly variable.
MYH3 nonframeshift deletion NM_002470: c.5263_5265delCTT g.chr17:10534948CCTT>C In one copy of the *MYH3* gene there is a sequence variant that deletes three bases and changes the amino acid sequence of the gene. This variant has not been seen in any dataset so far and appears to be unique.	This rare variant is of uncertain significance with respect to personal risk for Sheldon Hall and Freeman-Sheldon syndrome (Distal arthrogryposis types 2B and 2A, respectively) based on prediction of biochemical effect, evidence of sequence conservation, and rarity. Though this mutation has not been described in association with Sheldon Hall or Freeman-Sheldon syndromes, variants of this type have been implicated in these conditions.	Distal arthrogryposis syndromes are characterized by congenital contractures, predominantly in hands and feet. Freeman-Sheldon syndrome is the most severe, and is typically easily recognized due to striking orofacial contractures. When familial, Sheldon Hall and Freeman-Sheldon syndromes exhibit autosomal dominant inheritance, meaning that one or more improperly functioning copies of the *MYH3* gene must not function properly to cause the condition. The prevalence of *MYH3*-related distal arthrogryposis types 2B and 2A is unknown. Approximately 1:3000 children are born with arthrogryposis, but many cases are sporadic. Some individuals with genetic risk do not have clinical manifestations of the disease, and expressivity is highly variable.

Summary of Mendelian disease findings: Carrier status

Variant type	Variants
Carrier status	
Previously reported disease associated mutations	4
Rare, expected pathogenic variants with carrier status implications	1
Variants of uncertain significance	10

High level summary – Carrier status

Based on our current knowledge and analysis, we found four previously reported disease- associated mutations that have implications for carrier status for Wolman syndrome, alpha1-antitrypsin deficiency, LCHAD deficiency, and Canavan disease. We found one rare, expected pathogenic variant with carrier status implications for lissencephaly type 4. We found ten rare variants with uncertain carrier status implications in Mendelian disease-associated genes. Current genetic evidence is insufficient to confidently classify these variants as either disease associated or benign. Correlation with clinical and family history is advised.

Carrier status for Mendelian disease

Previously reported disease associated mutations

Genetic Variant	Implication for carrier status	Disease description
LIPA splicing variant NM_001127605: c.309G>A (p.Q103Q) g.chr10:90982268C>T In one copy of the *LIPA* gene there is a sequence variant that may affect normal splicing of exon 8. This variant has been seen in 0.13% of European subjects sequenced as part of the 1000Genomes study.	This variant may confer carrier status for Wolman disease or cholesteryl ester storage disease (CESD). It affects the last base of exon 8, thus affecting the normal splice donor site. Two variants at this same splice site in the *LIPA* gene have been implicated in both Wolman disease and several patients with CESD. This sequence variant appears to represent the specific mutation seen in 5 patients with Wolman disease and CESD from 3 families.[1-4]	Wolman disease is an early-onset fulminant disorder of infancy with massive infiltration of the liver, spleen, and other organs by macrophages filled with cholesteryl esters and triglycerides. Death occurs early in life. CESD is a milder, later-onset disorder with primary hepatic involvement by macrophages engorged with cholesteryl esters. This slowly progressive visceral disease has a very wide spectrum of involvement ranging from early onset with severe cirrhosis to later onset of more slowly progressive hepatic disease with survival into adulthood. Wolman disease and CESD are inherited in an autosomal recessive manner, meaning two copies of the *LIPA* gene must not function properly in order to cause the condition. Wolman disease is very rare, with an estimated birth prevalence of less than one in 100,000 live births. The prevalence of LIPA-related CESD is unknown. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.
SERPINA1 missense variant *(AATD Z allele)* NM_000295: c.1096 G>A (p.E366K) g.chr14:94844947C>T In one copy of the *SERPINA1* gene there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 2.0% of European subjects sequenced as part of the 1000 Genomes study and NHLBI exome sequencing project.	This particular variant represents the common deficiency allele, PI*Z, in the *SERPINA1* gene. Individuals with two copies of this allele may develop alpha1-antitrypsin deficiency (α1ATD, AATD).[5]	Alpha1-antitrypsin deficiency (α1ATD, AATD) is characterized by chronic obstructive pulmonary disease (COPD) in adults and liver disease in children and adults. COPD, specifically emphysema, is the most common manifestation of AATD. Smoking is the major factor influencing the course of COPD. AATD is inherited in an autosomal recessive manner, meaning two copies of the *SERPINA1* gene must not function properly in order to cause the condition. AATD has an estimated prevalence of 1: 5,000-7,000 individuals in North America and 1: 1,500-3,000 in Scandinavians. Carriers of this AATD Z allele do not commonly have clinical manifestations of disease; however, individuals who smoke and are carriers of this variant may have a slightly increased risk of COPD.
HADHA missense variant NM_000182: c.1267G>C (p.E423Q) g.chr2:26418053C>G In one copy of the *HADHA* gene there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 0.047%of European subjects sequenced as part of the 1000 Genomes study.	This variant may confer carrier status for trifunctional protein deficiency or LCHAD deficiency. This sequence change has been described in more than 40 patients with mainly isolated LCHAD deficiency in the literature, either in the presence of another disease-causing mutation or a second copy of this variant. It is thought to be one of the most common mutations implicated in LCHAD deficiency.[6-9]	LCHAD deficiency is characterized by early-onset cardiomyopathy, hypoglycemia, neuropathy, and pigmentary retinopathy, and sudden death. Biochemical studies have identified the LCHAD deficiency and trifunctional protein deficiency as two groups of disorders, which have similar clinical features but vary based on affected enzyme activities within mitochondrial trifunctional protein. Most patients in the literature represent cases of isolated LCHAD deficiency. These conditions are inherited in an autosomal recessive manner, meaning two copies of the *HADHA* gene must not function properly in order to cause LCHAD or trifunctional protein deficiency. LCHAD deficiency has an unknown prevalence. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.
ASPA missense variant NM_000049: c.914C>A (p.A305E) g.chr17:3402354C>A In one copy of the *ASPA* gene there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 0.023% of European subjects sequenced as part of the 1000 Genomes study.	This variant may confer carrier status for Canavan disease. This particular sequence change represents the most common mutation implicated in Canavan disease in individuals of non-Ashkenazi Jewish descent to date. It may account for up to 60% of known disease-causing mutations seen in patients of non-Ashkenazi Jewish descent, and approximately 1% of disease causing mutations in patients with Jewish ancestry.[10,11]	Canavan disease is a neurological disorder that affects the breakdown and use of aspartic acid. It is most commonly seen in neonatal/infantile form, which is characterized by macrocephaly, lack of head control, and developmental delays usually noted by age three to five months. As children get older hypotonia becomes severe and failure to achieve independent sitting, ambulation, or speech become apparent. Life expectancy is usually into the teens. Mild/juvenile Canavan disease is characterized by mild developmental delay that can go unrecognized. Canavan disease is inherited in an autosomal recessive manner, meaning two copies of the *ASPA* gene must not function properly in order to cause the condition. Canavan disease has an unknown prevalence in the non-Ashkenazi Jewish population. In individuals of Ashkenazi Jewish descent, the estimated prevalence of carriers for the condition ranges from 1:40 to 1:80. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.

[1-4] 1) Klima et al. J Clin Invest. 1993; 92:2713-8. 2) Maslen et al. J Inherit Metab Dis. 1995; 18: 620-3. 3) Muntoni et al. Hum. Genet. 1995; 95: 491-494. 4) Aslanidis et al. Genomics. 1996; 33:85-93.
[5] Brantly et al. *Am J Med*. 1988; 84:13–31.
[6-9] 6) IJlst Biochim Biophys Acta. 1994; 1215:347-50. 7) IJlst et al. J Clin Invest. 1996; 98:1028-33. 8) Tyni et al. J Pediatr. 1997; 130:67-76. 9) Boutron, et al. Mol Genet Metab 2011; 103:341-348.
[10] Elpeleg et al. J Inherit Metab Dis. 1999;22:531-4.
[11] Kaul et al. Am J Hum Genet. 1994;55:34-41.

Rare, expected pathogenic variants with potential carrier status implications

Genetic Variant	Implication for carrier status	Disease description
NDE1 nonsense variant NM_017668: c.145G>T (p.E49X) g.chr16:15761204G>T In one copy of the *NDE1* gene there is a sequence variant that causes an early stop in the amino acid sequence of the gene. This genetic variant has not been seen in any population dataset so far and appears to be unique.	This variant may confer carrier status for lissencephaly type 4. Though this particular mutation has not been described in association with lissencephaly type 4, highly disruptive variants of this type have been implicated in this condition.	Lissencephaly type 4 is characterized by lissencephaly, a brain malformation involving absence of normal convolutions (folds) in the cerebral cortex, severe brain atrophy, microcephaly, and intellectual disability. *NDE1*-related lissencephaly is inherited in an autosomal recessive manner, meaning two copies of the *NDE1* gene must not function properly to cause the condition. *NDE1*-related lissencephaly has only been reported in four families in the literature to date. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.

Genetic variants of uncertain carrier status significance

Genetic Variant	Implication for carrier status	Disease description
PLOD1 missense variant NM_000302: c.1534C>T (p.R512C) g.chr1:12025600C>T In one copy of the *PLOD1* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 0.55% of European subjects sequenced as part of the NHLBI exome sequencing project and 1% of European subjects as part of the 1000 Genomes study.	This rare variant is of uncertain significance with respect to carrier status for Ehlers Danlos syndrome, kyphoscoliotic form (EDS, kyphoscoliotic form) based on evidence of sequence conservation, prediction of biochemical effect, and rarity. While this particular mutation has not been described in association with this form of EDS, variants of this type have been implicated in this condition.	Ehlers-Danlos syndrome (EDS), kyphoscoliotic form (previously known as EDS VI) is a generalized connective tissue disorder characterized by friable, hyperextensible skin, thin scars, and easy bruising; generalized joint laxity; severe muscular hypotonia at birth; progressive scoliosis, present at birth or within the first year of life; and scleral fragility with increased risk of rupture of the globe. EDS kyphoscoliotic form is inherited in an autosomal recessive manner, meaning two copies of the *PLOD1* gene must not function properly to cause the condition. EDS, kyphoscoliotic form is rare; the exact prevalence is unknown. The estimated incidence of EDS is approximately 1:100,000 live births. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.
HSPG2 missense variant NM_005529: c.1033G>A (p.D345N) g.chr1:22213753C>T In one copy of the *HSPG2* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has not been seen any control datasets and appears to be unique.	This novel variant is of uncertain significance with respect to carrier status for Schwartz-Jampel syndrome (SJS) type 1 based on evidence of sequence conservation, prediction of biochemical effect, and rarity. While this particular mutation has not been described in association with SLS type 1, variants of this type have been implicated in the disorder.	Schwartz-Jampel syndrome (SJS) type 1 is characterized by multiple manifestations including short stature, abnormal facies, pectus carinatum, hernia, and myotonic dystrophy. SJS type 1 is inherited in an autosomal recessive manner, meaning both copies of the *HSPG2* gene must not function properly to cause the condition. SJS is very rare, with unknown prevalence. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.
DLD missense variant NM_000108: c.100A>G (p.T34A) g.chr7:107533705A>G In one copy of the *DLD* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 0.16% of European subjects sequenced as part of the NHLBI exome sequencing project.	This rare variant is of uncertain significance with respect to carrier status for E3-deficient maple syrup urine disease based on evidence of sequence conservation, prediction of biochemical effect, and rarity. Though this particular mutation has not been described in association with E3-deficient maple syrup urine disease, variants of this type have been implicated in the disorder.	E3-deficient maple syrup urine disease is distinct from classic maple syrup urine disease, where affected infants typically have hypotonia, developmental delay, dystonia/chorea, and Leigh-type encephalopathy. In most cases, the disorder is lethal in infants. This condition is inherited in an autosomal recessive manner, meaning that both copies of the *DLD* gene must not function properly to cause the condition. The prevalence E3-deficient maple syrup urine disease is unknown. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.
PLEC missense variant NM_201384: c.5053C>T (p.R1685W) g.chr8:144999044G>A In one copy of the *PLEC* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has not been seen in any control datasets and appears to be unique.	This novel variant is of uncertain significance with respect to carrier status for Epidermolysis Bullosa with Pyloric Atresia (EB-PA) and Epidermolysis Bullosa with Muscular Dystrophy (EB-MD) based on evidence of sequence conservation, prediction of biochemical effect, and rarity. Variants of this type have been implicated in both Epidermolysis Bullosa with Pyloric Atresia and Epidermolysis Bullosa with Muscular Dystrophy. Mutations in this gene have also been implicated in another condition, Limb-Girdle Muscular Dystrophy (LGMD), but generally this condition is not caused by variants of this type.	Epidermolysis Bullosa-Pyloric Atresia is characterized by fragility of the skin and mucous membranes, congenital pyloric atresia (absence of normal passage between stomach and small intestine), and ureteral and kidney abnormalities. Epidermolysis Bullosa-Muscular Dystrophy is characterized by similar clinical manifestations but without pyloric atresia and with muscular dystrophy of variable age of onset. This condition is inherited in an autosomal recessive manner, meaning both copies of the *PLEC* gene must not function properly to cause disease. Epidermolysis Bullosa-Pyloric Atresia is rare and its prevalence and incidence have not been determined. Approximately 50 cases of Epidermolysis Bullosa-Muscular Dystrophy have been reported worldwide. Current knowledge indicates that carriers for these conditions do not have clinical manifestations of the disease.
PCDH15 missense variant NM_001142765: c.3100C>A (p.L1034I) g.chr10:55698635G>T In one copy of the *PCDH15* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has not been seen in any control datasets and appears to be unique.	This novel variant is of uncertain significance with respect to carrier status for autosomal recessive nonsyndromic hearing loss (DFNB23) based on evidence of sequence conservation, prediction of biochemical effect, and rarity. While this particular mutation has not been described in association with DFNB23, variants of this type have been implicated in the disorder. Mutations in this gene have also been implicated in another condition, Usher syndrome type F, but generally this condition is not caused by variants of this type.	Nonsyndromic hearing loss is characterized by hereditary hearing loss, not associated with visible abnormalities of the external ear or any related medical issues. *PCDH15*-associated HL is called DFNB23 and is associated with prelingual onset, and stable severe to profound hearing loss. *PCDH15*-associated hearing loss is inherited in an autosomal recessive manner, meaning both copies of the *PCDH15* gene must not function properly to cause the condition. The prevalence of *PCDH15*-associated hearing loss is unknown. One of every 500 newborns has bilateral permanent sensorineural hearing loss ≥40 dB; by adolescence, prevalence increases to 3.5 per 1000. Current knowledge indicates that carriers for these conditions do not have clinical manifestations of the disease.

SLC39A13 missense variant NM_152264: c.191G>A (p.R64Q) g.chr11:47431836G>A In one copy of the *SLC39A13* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 0.012% of European subjects sequenced as part of the NHLBI exome sequencing project.	This rare variant is of uncertain significance with respect to carrier status for Spondylocheirodysplasia, EDS-like (SCD-EDS) based on evidence of sequence conservation, prediction of biochemical effect, and rarity. While this particular mutation has not been described in association with SCD-EDS, variants of this type have been implicated in the disorder.	Spondylocheirodysplasia, EDS-like (SCD-EDS) is characterized by clinical features that include typical EDS characteristics of hyperelastic, thin, burisable skin, hypermobile joints, and a generalized skeletal dysplasia and abnormalities of the hands. SCD-EDS is inherited in an autosomal recessive manner, meaning both copies of the *SLC39A13* gene must not function properly to cause the condition. SCD-EDS has only been reported in a small number of families worldwide. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.
SEC23B missense variant NM_006363: c.1198T>C (p.F400L) g.chr20:18511412C>T In one copy of the *SEC23B* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 0.26% of European subjects sequenced as part of the 1000 Genomes study, and 0.093% of European subjects sequenced as part of the NHLBI exome sequencing project.	This rare variant is of uncertain significance with respect to carrier status for Congenital dyserythropoietic anemia type II (CDA II) based on evidence of sequence conservation, prediction of biochemical effect, and rarity. While this particular mutation has not been described in association with congenital disorder of glycosylation, variants of this type have been implicated in the disorder.	Congenital dyserythropoietic anemia type II (CDA II) is the most common CDA, characterized by mild to severe anemia, jaundice, and splenomegaly, which is observed in 50%-60% of affected individuals. Beyond age 20 years most affected individuals develop iron overload. The diagnosis of CDA II requires evidence of congenital anemia, ineffective erythropoiesis, and typical bone marrow findings with binuclearity in 10%-50% of erythroblasts. CDA II is inherited in an autosomal recessive manner, meaning both copies of the *SEC23B* gene must not function properly to cause the condition. CDA type II is the most common form of congenital dyserythropoietic anemia, though it has an unknown prevalence. CDA as a whole has been reported in several hundred cases worldwide. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.
CDT1 missense variant NM_030928: c.248C>T (p.P83L) g.chr16:88870972C>T In one copy of the *CDT1* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 1.0% of European subjects sequenced as part of the 1000 Genomes study, and 0.20% of European subjects sequenced as part of the NHLBI exome sequencing project.	This rare variant is of uncertain significance with respect to carrier status for Meir-Gorlin syndrome type 4 (MGORS4) based on evidence of sequence conservation, prediction of biochemical effect, and rarity. While this particular mutation has not been described in association with congenital disorder of glycosylation, variants of this type have been implicated in the disorder.	Meier-Gorlin syndrome is characterized by severe intrauterine and postnatal growth retardation, microcephaly, bilateral microtia, and aplasia or hypoplasia of the patellae. Meier-Gorlin syndrome is also referred to as ear, patella and short-stature syndrome. There are many genes implicated in this condition. *CDT1*-related Meier-Gorlin syndrome is inherited in an autosomal recessive manner, meaning both copies of the *CDT1* gene must not function properly to cause the condition. *CDT1*-related Meier-Gorlin syndrome has only been reported in 53 cases in the literature. Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.
PKLR missense variant NM_181871: c.1423G>A (p.V475I) g.chr1:155261649C>T In one copy of the *PKLR* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 1.0% of European subjects sequenced as part of the 1000 Genomes study, and 0.51% of European subjects sequenced as part of the NHLBI exome sequencing project.	This rare variant is of uncertain significance with respect to carrier status for pyruvate kinase (PK) deficiency. This particular variant has been described in two patients exhibiting PK deficiency. However, the presented evidence does not clearly implicate this particular variant in the condition. One case was a 9-year-old patient from Palma, Mallorca with hereditary spherocytosis, moderate hemolytic anemia, and mild pyruvate kinase deficiency, where the evidence presented supporting its role in disease was weak. The second case was in a French patient in which few details were presented. In this case, the patient presented with neonatal jaundice and was identified with this particular variant and another variant of this type in the *PKLR* gene. [1,2]	Pyruvate kinase deficiency is the most frequent cause of congenital nonspherocytic hemolytic anemia. PK deficiency essentially differs from hereditary spherocytosis by the absence of circulating spherocytes and a more severe chronic hemolytic anemia. PK deficiency is inherited in an autosomal recessive manner, meaning both copies of the *PKLR* gene must not function properly to cause the condition. PK deficiency has an estimated prevalence of 1:20,000. Current knowledge indicates that some carriers for this condition may have mild manifestations of the disease.
PKHD1 missense variant NM_138694: c.5125C>T (p.L1709F) g.chr6: 51889483G>A In one copy of the *PKLR* gene, there is a sequence variant that changes the amino acid sequence of the gene. This variant has been seen in 0.49% of European subjects sequenced as part of the NHLBI exome sequencing project.	This rare variant is of uncertain significance with respect to carrier status for autosomal recessive polycystic kidney disease (ARPKD) based on evidence of sequence conservation, prediction of biochemical effect, and rarity. This particular variant has been identified in two unrelated patients with ARPKD of South African and European descent. In both individuals another variant of this type was also found. However, there are also several studies in the literature that have described this particular variant in patients with ARPKD, but authors classified it as a polymorphism that is unlikely to confer carrier status on the basis of the frequency in healthy control individuals and ARKPD patients.[3-6] Given the conflicting lines of evidence in the literature, this variant is currently considered to be of unknown significance in terms of carrier status for polycystic kidney disease.	Polycystic kidney disease is a condition that affects the kidneys and other organs. The majority of individuals with autosomal recessive polycystic kidney disease (ARPKD) present in the neonatal period with enlarged echogenic kidneys. At initial presentation, approximately half of infants have liver abnormalities, including hepatomegaly, dilated intrahepatic (and occasionally extrahepatic) biliary ducts, and mildly increased echogenicity. ARPKD is inherited in an autosomal recessive manner, meaning both copies of the *PKHD1* gene must not function properly to cause the condition. ARPKD has an estimated birth prevalence ranging from 1:10,000 to 1:40,000. The carrier frequency for a *PKHD1* mutation in the general population is estimated to be 1:70. [7] Current knowledge indicates that carriers for this condition do not have clinical manifestations of the disease.

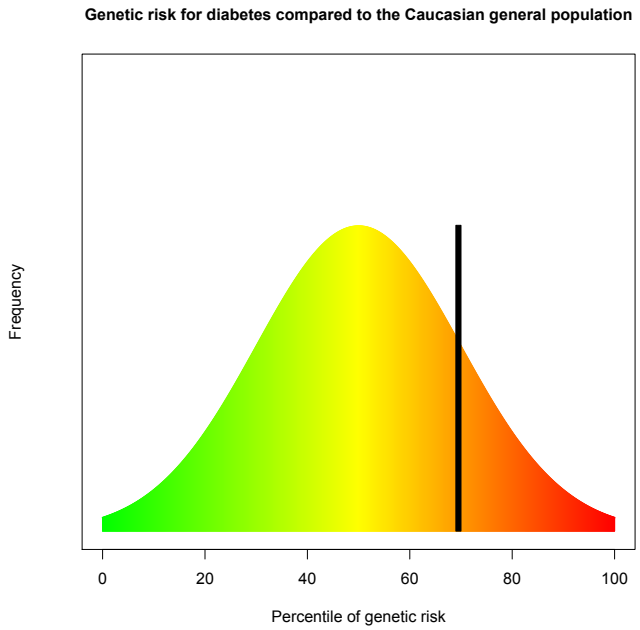
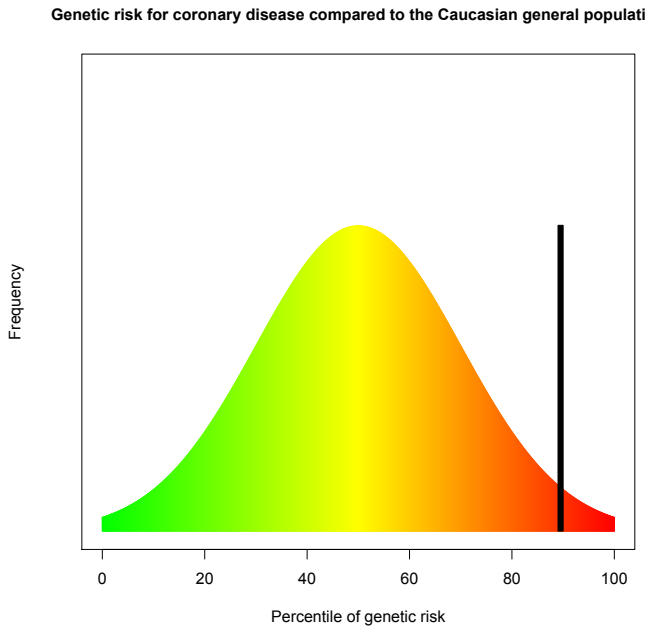
[1] Zarza et al. Haematologica. 2000; 85:227-32.
[2] Pissard et al. Br J Haematol. 2006;133:683-9.
[3-6] 3) Furu et al. J Am Soc Nephrol. 2003;14:2004-14. 4) Rossetti et al. Kidney Int. 2003;64:391-403. 5) Losekoot et al. Hum Genet. 2005;118:185-206. 6) Gunay-Aygun et al. Mol Genet Metab. 2010;99:160-73.
[7] Zerres et al. Mol Genet Metab. 2010;99:160-73.

Cardiometabolic disease risk

Technical summary

We scanned the genome for 90 genetic risk markers for coronary artery disease and 96 genetic risk markers for diabetes that have been established in Caucasian populations. Both diseases have complex genetic and environmental contributions. There is evidence that these genetic risk markers may augment the predictive power of traditional clinical assessment for these two diseases. However, in general, the proportion of common disease risk explained by genetic variants discovered thus far is small. It is important to remember that *other genetic and environmental factors may also influence risk for common disease.*

Of the genetic risk markers that have strong evidence to support coronary artery disease or diabetes risk associations, we were able to confidently sequence 100%. A summary of genetic risk according to these markers is provided below.

Diabetes	Coronary artery disease
We tested for a total of 96 possible genetic risk factors for diabetes. Out of these 96 we found 54 that are associated with higher risk. This genetic profile is in the 69th percentile for risk when compared to the Caucasian population. This means that this genetic risk profile is more favorable than 31% of the Caucasian population, and less favorable than 69% of the Caucasian population.	We tested for a total of 90 possible genetic risk factors for coronary artery disease. Out of these 90, we found 42 that are associated with higher risk. This genetic profile is in the 89th percentile for risk when compared to the Caucasian population. This means that this genetic risk profile is more favorable than 11% of the Caucasian population, and less favorable than 89% of the Caucasian population.
	

Guideline-informed genetic variants influencing drug response

Technical summary

We scanned the genome for 555 genetic changes that have been found to affect response to 162 commonly used medications. Though there are many described genetic associations with drug response that have unclear clinical implications, the genetic associations with three commonly used medications in cardiovascular medicine, warfarin, clopidogrel, and simvastatin, may impact prescribing recommendations. It is important to remember that *other genetic and clinical factors may also influence risk for medication-related adverse events.*

Of the 555 genetic variants associated with drug response, we were able to confidently sequence 548 (98.7%). Using this genetic information, we were able to determine drug response predictions and usage recommendations according to guidelines set forth by the Clinical Pharmacogenomics Implementation Consortium for warfarin, clopidogrel, codeine, thiopurines, and simvastatin.[1-5] A summary of these drug response predictions and usage recommendations is provided below.

[1] Relling MV et al. Clin Pharmacol Ther 2013; 93:324-5.
[2] Wilke RA et al. Clin Pharmacol Ther 2012; 92:112-7.
[3] Crews KR et al. Clin Pharmacol Ther 2012;91:321-6.
[4] Johnson JA et al. Clin Pharmacol Ther 2011;90:625-9.
[5] Scott SA et al. Clin Pharmacol Ther 2011;90:328-32.

Genotype	Implication for drug response	Drug description
SLCO1B1: rs4149056 TT (Homozygous functional allele) In the *SLCO1B1* gene, there are two copies of the most common allele that is associated with normal gene function. Individuals with one or more improperly functioning copies of this gene may be at increased risk of simvastatin-related myopathy when taking this drug.	Given the available genetic evidence for simvastatin, which includes *SLCO1B1* genetic status, there may be no increased risk of simvastatin-related myopathy. **For someone with this specific genotype, the recommendation is to use the standard label-recommended dose and administration for simvastatin.**	Simvastatin is one of the most commonly prescribed cholesterol- lowering agents. It is used to lower LDL (bad cholesterol) and raise HDL (good cholesterol), and to prevent stroke and heart attack. Simvastatin is partially metabolized in the liver. Inadequate metabolism of this drug may lead to muscle injury (myopathy). Genetic factors, including variation in the gene *SLC01B1*, can influence an individual's ability to metabolize simvastatin and therefore the risk of myopathy. There is some less well-established evidence for similar genetic effects on other drugs in the statin class.
CYP2C19: *1/*17 In the *CYP2C19* gene, there is one copy of an allele with possible reduced metabolism and one copy of an allele with increased metabolism. Individuals with one or more improperly functioning copies of this gene may be at increased risk of drug inefficacy and cardiovascular events such as heart attack or stroke when taking this drug.	Given the available genetic evidence for clopidogrel metabolism, which includes *CYP2C19* genetic status, platelet inhibition with clopidogrel use may be higher than someone with two copies of the most common allele. **For someone with this specific genotype, the recommendation is to use the standard label-recommended dose and administration for clopidogrel.**	Clopidogrel is one of the most commonly prescribed clot-inhibiting agents and is used to prevent and treat stroke and heart attack and to prevent complications related to coronary stent implantation. Clopidogrel requires metabolism in the liver for inhibition of clotting. There is high inter-individual variability in metabolism of this drug that affects both the therapeutic effect and the risk for adverse events when taking this drug. Genetic factors, including common variation in the gene *CYP2C19*, can influence an individual's ability to metabolize clopidogrel and therefore its therapeutic effect.
CYP2C9: *1/*1 *VKORC1*: rs9923231 TC (Heterozygous nonfunctional allele) In the *CYP2C9* gene, there are two copies of the most common allele. There is one copy of an allele of *VKORC1* that metabolizes warfarin at lower levels. Individuals with one or more improperly functioning copies of both of these *CYP2C9* and *VKORC1*, or two improperly functioning copies of one of these genes, may be at increased risk of bleeding or cardiovascular events such as heart attack or stroke when taking this drug.	Given the available genetic evidence for warfarin metabolism, which includes *VKORC1* and *CYP2C9* genetic status, the required dose of warfarin is likely to be near the standard dose. **For someone with this specific genotype, the recommendations are for the standard dose, around 5-7 mg/day.**	Warfarin is the most commonly prescribed blood-thinning agent and is used to prevent blood clots, strokes and heart attacks. This drug is difficult to dose due to high inter-individual variability and a narrow therapeutic index. Genetic factors, including variation in the genes *CYP2C9* and *VKORC1*, can influence an individual's ability to metabolize warfarin.
CYP2D6: *1/*4 In the *CYP2D6* gene there is one copy of the allele associated with full function of codeine metabolism and one copy of an allele with reduced function. Individuals with two improperly functioning alleles may have reduced formation of morphine from codeine when compared to the individuals with one or more copies of the fully functional allele, and individuals with more than two copies of the fully functional allele may have increased formation of morphine.	Given the available genetic evidence for codeine metabolism, which includes *CYP2D6* genetic status, the required dose of codeine is likely to be near the standard dose. **For someone with this specific genotype, the recommendation is to use the standard label recommended dose and administration, 15-50 mg every four hours as needed for pain.**	Codeine is a commonly prescribed drug for the treatment of pain (analgesia) and requires metabolism in the liver to morphine, a strong opioid analgesic. Rapid metabolism of this drug can lead to morphine toxicity, whereas slow metabolism can lead to limited therapeutic effect. Genetic factors, including variation in the gene *CYP2D6*, can influence an individual's ability to metabolize codeine to morphine.
TPMT: *1/*1 In the *TPMT* gene there are two copies of the most common allele that is associated with normal gene function. Individuals with one or more improperly functioning alleles may have reduced metabolism of thiopurines, increased thioguanine nucleolide levels, and greater bone marrow suppression at standard doses.	Given the available genetic evidence for thiopurine metabolism, which includes *TPMT* genetic status, the level of thioguanine nucleolides and degree of bone marrow suppression with standard doses are likely to be typical. **For someone with this specific genotype, the recommendation is to start with normal starting dose and adjust doses based on disease specific guidelines. Adjust doses of thiopurine (and of any other myelosuppressive therapy) without any special emphasis on thiopurines compared to other agents. Allow 2 weeks to reach steady state after each dose adjustment.**	Thiopurines such as azathioprine, mercaptopurine, and thioguanine are commonly used to treat autoimmune conditions, as immunosuppresants for organ transplant recipients, and for treatment of cancer. Thiopurines may be difficult to dose due to high inter-individual variability and narrow therapeutic index. Genetic factors, including variation in the gene *TPMT*, can influence an individual's ability to metabolize thiopurines and therefore its risk of toxicity.

Appendix A – databases used for genome analysis

Databases used for genome analysis:

Allele frequency estimation NHLBI Exome Sequencing Project 6500 (all) NHLBI Exome Sequencing Project 6500 (European Americans) NHLBI Exome Sequencing Project 6500 (African Americans) HapMap 2 and 3 African ancestry in Southwest USA HapMap 2 and 3 Utah residents with Northern and Western European ancestry from the CEPH collection HapMap 2 and 3 Han Chinese in Beijing, China HapMap 2 and 3 Chinese in Metropolitan Denver, Colorado HapMap 2 and 3 Gujarati Indians in Houston, Texas HapMap 2 and 3 Japanese in Tokyo, Japan HapMap 2 and 3 Luhya in Webuye, Kenya HapMap 2 and 3 Mexican ancestry in Los Angeles, California HapMap 2 and 3 Maasai in Kinyawa, Kenya HapMap 2 and 3 Toscani in Italia HapMap 2 and 3 Yoruba in Ibadan, Nigeria 1000 genomes project pilot 1, 2010 November release, all subjects 1000 genomes project phase 1, 2011 May release, all subjects 1000 genomes project phase 1, 2012 February release, all subjects 1000 genomes project phase 1, 2012 April release, all subjects 1000 genomes project phase 1, 2012 April release, European ancestry 1000 genomes project phase 1, 2012 April release, East Asian ancestry 1000 genomes project phase 1, 2012 April release, West African ancestry 1000 genomes project phase 1, 2012 April release, American ancestry Complete Genomics public panel (cg46) Complete Genomics diversity panel (cg69) NCBI dbSNP database, version 135 **Gene effect prediction** UCSC known gene NCBI RefSeq Gencode	**Functional effect prediction** LRT SIFT PolyPhen2 Mutation Taster dbNSFP Segmental duplication database Database of Genomic Variants **Evolutionary sequence comparison** PhastCons PhyloP GERP SIFT Polyphen2 **Medical effect prediction** Human Gene Mutation Database NCBI ClinVar NHGRI Genome Wide Association Study catalog VariMed PharmGKB CARDIOGRAM C4D TAICHI

Appendix B – criteria for reporting of Mendelian variants

Criteria for reporting:

- **Reported disease-associated mutations:**
  - Mutations convincingly segregating with Mendelian disease in one or more families and absent or observed very infrequently* in ethnically matched population controls.
  - Mutations observed in multiple unrelated probands with Mendelian disease, and absent or observed very infrequently* in ethnically matched population controls.
  - Mutations observed in single probands with Mendelian disease, with functional evidence of pathogenicity, and absent or observed very infrequently* in ethnically matched population controls.
- **Rare, expected pathogenic variants with potential implications for personal disease risk or carrier status:**
  - Rare nonsense, stop loss, splice-disrupting, or frameshift insertion/deletion variants affecting the majority of protein coding transcripts of genes associated with Mendelian disease.
- **Genetic variants of uncertain personal disease risk or carrier status significance:**
  - Rare missense variants with evolutionary conservation or biochemical prediction support for predicted pathogenicity, rare frameshift variants, and variants of all types previously reported as disease causing in which primary literature reports of pathogenicity provide conflicting or incomplete evidence of causality. These variants must occur in Mendelian disease genes in which there is evidence that variants of the type discovered in study

** As defined by an allele frequency consistent with mode of inheritance and population prevalence of disease.*

Appendix C – Genetics glossary

Allele: "Allele" is the term used to describe specific forms or versions of a gene.

Allele frequency: The proportion of chromosomes in a population harboring a specific allele. "Minor allele frequency" refers to the frequency of the less common allele at a biallelic location in the genome.

Chromosome: Linear or circular DNA strand.

Codon: A sequence of three nucleotides, which together form a unit of genetic code in a DNA molecule.

Deletion: A DNA sequence variant in which one or more DNA bases are deleted.

DNA: Deoxyribonucleic acid. A DNA base, or nucleotide, is the primary structure that makes up a DNA molecule. There are four different types of DNA bases; each represented by one of four letters (A, C, G, T).

Dominant inheritance: Autosomal dominant inheritance means that one or more copies of a disease-associated gene must not function properly to cause the condition.

Frameshift: A sequence variant that results in a shift in the normal reading frame of the gene and thus alters the protein.

Gene: A segment of DNA that provides specific instructions on how to make proteins and other important biologic substances for the body. The human genome contains ~22,000 distinct genes.

Genome: The genome is the full complement of DNA carried by an individual and the collection of proteins required to read, maintain, and give DNA its structure.

Genotype: The combination of alleles at one genomic location in an individual.

Germline: Gives rise to gametes (eggs, sperm), which contain the genetic material that may be inherited.

Insertion: A DNA sequence variant in which one or more bases are inserted.

Locus: A specific genomic location.

Phenotype: Physical characteristics of an individual, for instance expression of a disease or other trait.

Read depth: The number of independent times a location in the genome has been sequenced.

Reading frame: The frame of reference for translating the DNA sequence to protein.

Recessive inheritance: Autosomal recessive inheritance means that both copies of a disease-associated gene must not function properly to cause the condition.

Risk allele: The allele that has been shown to be associated with risk of a disease.

Sequencing: A method for determining the precise order of bases in a DNA molecule.

SNP: Single nucleotide polymorphism. A site in the genome where a single DNA base has been found to be variable in human populations.

Substitution: A DNA sequence variant in which one base is substituted for another.